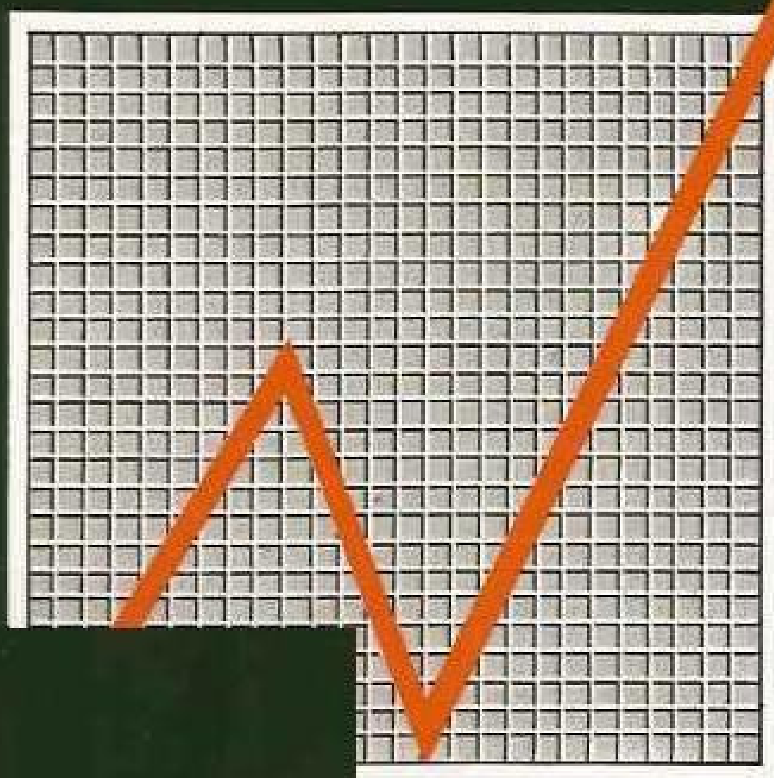
JAIRO SIMON DA FONSECA
GILBERTO DE ANDRADE MARTINS

CURSO DE ESTATÍSTICA



atlas

CURSO DE ESTATÍSTICA

JAIRO SIMON DA FONSECA
GILBERTO DE ANDRADE MARTINS

CURSO DE ESTATÍSTICA

6ª Edição

SÃO PAULO
EDITORA ATLAS S.A. – 2010



2. ed. 1975; 3. ed. 1979; 4. ed. 1993; 5. ed. 1994;
6. ed. 1996; 13. reimpressão 2010

Capa: Ronaldo Rizzulli

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Fonseca, Jairo Simon da

Curso de estatística / Jairo Simon da Fonseca, Gilberto de Andrade Martins. – 6. ed. – 13. reimpr. – São Paulo : Atlas, 2010.

ISBN 978-85-224-1471-0

1. Estatística 2. Estatística matemática 3. Estatística – Problemas, exercícios etc. 4. Probabilidades I. Martins, Gilberto de Andrade, 1947 – II. Título.

92-0830 CDD-519.5

Índices para catálogo sistemático:

1. Estatística 519.5
2. Estatística matemática 519.5

Depósito legal na Biblioteca Nacional conforme Decreto nº 1.825, de 20 de dezembro de 1907.

Impresso no Brasil/*Printed in Brazil*



Editora Atlas S.A.
Rua Conselheiro Nébias, 1384 (Campos Elísios)
01203-904 São Paulo (SP)
Tel.: (0_ _11) 3357-9144 (PABX)
www.EditoraAtlas.com.br

SUMÁRIO

PREFÁCIO

Decorridos 15 anos de seu lançamento, estamos apresentando uma nova e revisada edição do Curso de Estatística.

Aproveitando as críticas e sugestões enviadas por colegas de todo o país, bem como nossas experiências na condução de cursos de Estatística, cremos poder oferecer um livro-texto que atenda professores e alunos de um primeiro curso universitário de Probabilidade e Estatística.

Nossa preocupação foi a de produzir um texto claro e compreensível sobre probabilidades e estatística, mantendo características eminentemente didáticas.

Esta nova versão está fundamentada no aumento da quantidade de exercícios resolvidos e propostos e no tratamento de outros métodos estatísticos que não fizeram parte das edições anteriores.

Cada capítulo inicia-se com uma exposição objetiva do assunto indicado, solução de exemplos e proposição de exercícios cujas respostas se encontram no final do livro.

Os quatro primeiros capítulos dedicam-se ao estudo das probabilidades, variáveis aleatórias, modelos de distribuição discretas e contínuas de probabilidades.

A Estatística Descritiva é detalhada no Capítulo 5, enquanto os Métodos sobre Inferência são apresentados nos Capítulos 6, 7, 8 e 9.

O Capítulo 10 é dedicado ao tratamento de testes e provas não paramétricas. Assunto, particularmente, interessante aos estudante e pesquisadores das áreas de humanidades.

A Análise da Variância: hipóteses, fundamentos e principais modelos são explicados no Capítulo 11.

Para auxílio ao leitor são apresentadas, no final do livro, as principais tabelas estatísticas, que também constituem anexo útil para consultas em aulas de exercícios e avaliações.

Os Autores

1

CÁLCULO DAS PROBABILIDADES

1.1 INTRODUÇÃO

Todas as vezes que se estudam fenômenos de observação, cumpre-se distinguir o próprio fenômeno e o modelo matemático (determinístico ou probabilístico) que melhor o explique.

Os fenômenos estudados pela Estatística são fenômenos cujo resultado, mesmo em condições normais de experimentação variam de uma observação para outra, dificultando dessa maneira a previsão de um resultado futuro.

Para a explicação desses fenômenos – fenômenos aleatórios – adota-se um modelo matemático probabilístico. Neste caso, o modelo utilizado será o CÁLCULO DAS PROBABILIDADES.

1.2 CARACTERIZAÇÃO DE UM EXPERIMENTO ALEATÓRIO

A fim de se entender melhor a caracterização desses experimentos, convém observar o que há de comum nos seguintes experimentos:

E_1: Retirar uma carta de um baralho com 52 cartas e observar seu "naipe".

E_2: Jogar uma moeda 10 vezes e observar o número de coroas obtidas.

E_3: Retirar com ou sem reposição, bolas de uma urna que contém 5 bolas brancas e 6 pretas.

E_4: Jogar um dado e observar o número mostrado na face de cima.

E_5: Contar o número de peças defeituosas da produção diária da máquina A.

A análise desses experimentos revela:

a) Cada experimento poderá ser repetido indefinidamente sob as mesmas condições ;

b) Não se conhece um particular valor do experimento "a priori", porém pode-se descrever todos os possíveis resultados – as possibilidades;

c) Quando o experimento for repetido um grande número de vezes surgirá uma regularidade, isto é, haverá uma estabilidade da fração $f = \frac{r}{n}$ (freqüência relativa), em que n é o número de repetições e r o número de sucessos de um particular resultado estabelecido antes da realização.

Assim,

f
1
0
n

Como vai se ver mais adiante, a característica (c) é de fundamental importância para a avaliação da probabilidade de um certo evento.

1.3 ESPAÇO AMOSTRAL

Definição: Para cada experimento aleatório E, define-se Espaço Amostral S o conjunto de todos os possíveis resultados desse experimento.

Exemplos:

a) Considere-se o experimento
E = jogar um dado e observar o nº da face de cima.
então, $S = \{1, 2, 3, 4, 5, 6\}$

b) Seja E: jogar duas moedas e observar o resultado.
então, $S = \{(c, c), (c, k), (k, c), (k, k)\}$
em que c = cara e k = coroa.

Observe que sendo S um conjunto, poderá ser finito ou infinito.

Neste capítulo serão considerados apenas conjuntos finitos.

1.4 EVENTO

Definição: evento é um conjunto de resultados do experimento, em termos de conjuntos, é um subconjunto de S. Em particular, S e ϕ (conjunto vazio) são eventos, S é dito o evento certo e ϕ o evento impossível.

Usando as operações com conjuntos, podem-se formar novos eventos. Assim:

I) $A \cup B \rightarrow$ é o evento que ocorre se A ocorre ou B ocorre ou ambos ocorrem;

II) $A \cap B \rightarrow$ é o evento que ocorre se A e B ocorrem;

III) $\bar{A} \rightarrow$ é o evento que ocorre se A não ocorre.

Exemplos:

a) Seja o experimento E: jogar três moedas e observar os resultados:

Então: $S = \{(c, c, c), (c, c, k), (k, c, c), (c, k, c), (k, k, k), (k, k, c), (k, c, k), (c, k, k)\}$

Seja A o evento: ocorrer pelo menos 2 caras.

Então, $A = \{(c, c, c), (c, c, k), (c, k, c), (k, c, c)\}$

b) Seja o experimento E: lançar um dado e observar o número de cima.

Então $S = \{1, 2, 3, 4, 5, 6\}$

Seja B o evento: ocorrer múltiplo de 2.

Então, $B = \{2, 4, 6\}$

Sendo S espaço amostral finito, com n elementos pode-se verificar que 2^n fornece o número total de eventos extraídos de S.

1.5 EVENTOS MUTUAMENTE EXCLUSIVOS

Dois eventos A e B são denominados mutuamente exclusivos, se eles não puderem ocorrer simultaneamente, isto é, $A \cap B = \phi$.

Exemplo: E: jogar um dado e observar o resultado.

$S = \{1, 2, 3, 4, 5, 6\}$

Sejam os eventos: A = ocorrer nº par, e

B = ocorrer nº ímpar.

Então, $A = \{2, 4, 6\}$ e $B = \{1, 3, 5\}$, $A \cap B = \phi$

A e B são mutuamente exclusivos, pois a ocorrência de um número par e ímpar não pode ser verificada como decorrência da mesma experiência.

1.6 DEFINIÇÃO DE PROBABILIDADE

Dado um experimento aleatório E e S o espaço amostral, probabilidade de um evento A – $P(A)$ – é uma função definida em S que associa a cada evento um número real, satisfazendo os seguintes axiomas:

I) $0 \leq P(A) \leq 1$

II) $P(S) = 1$

III) Se A e B forem eventos mutuamente exclusivos, $(A \cap B = \phi)$, então $P(A \cup B) = P(A) + P(B)$

1.7 PRINCIPAIS TEOREMAS

1. Se ϕ é o conjunto vazio, então $P(\phi) = 0$

Demonstração:

Seja A um evento qualquer.

A e ϕ são disjuntos, pois $A \cap \phi = \phi$

$P(A \cup \phi) = P(A) + P(\phi)$ por III.

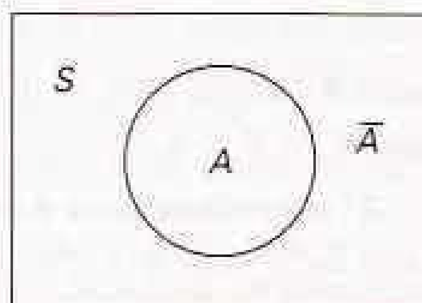
$P(A) = P(A) + P(\phi)$ pois $A \cup \phi = A$

Portanto $P(\phi) = 0$

2. Se $\bar{A}$ é o complemento do evento A, então $P(\bar{A}) = 1 - P(A)$

Demonstrações:

Pode-se escrever $S = A \cup \bar{A}$. (Veja o diagrama).



ora, $A \cap \bar{A} = \phi$ (são mutuamente exclusivos)

$P(A \cup \bar{A}) = P(A) + P(\bar{A})$

$P(S) = P(A) + P(\bar{A})$

$1 = P(A) + P(\bar{A})$ por II.

Logo: $P(\bar{A}) = 1 - P(A)$

3. Se $A \subset B$, então $P(A) \leq P(B)$

Demonstração:

Pode-se escrever $B = A \cup (\bar{A} \cap B)$

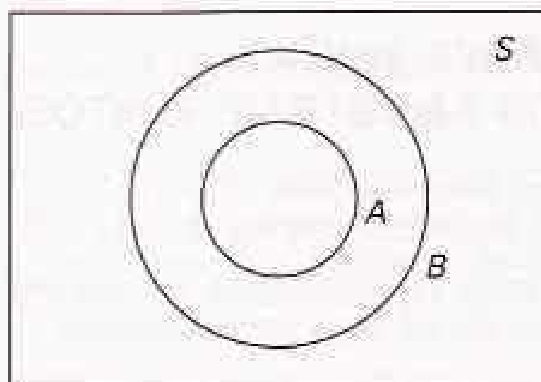
ora, A e $(\bar{A} \cap B)$ são mutuamente exclusivos.

Logo, $P(B) = P(A) + P(\bar{A} \cap B)$

$P(\bar{A} \cap B) = P(B) - P(A)$

$P(B) - P(A) \geq 0$ por I

Portanto, $P(A) \leq P(B)$.



4. Teorema da soma: Se A e B são dois eventos quaisquer, então:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Demonstração:

a) A e B são mutuamente exclusivos $P(A \cap B) = 0$ recai-se no axioma III.

b) Para o caso em que $A \cap B \neq \phi$.

Os eventos A e $(\bar{A} \cap B)$ são mutuamente exclusivos. Logo, pelo axioma III: $P[A \cup (\bar{A} \cap B)] = P(A \cup B) = P(A) + P(\bar{A} \cap B)$. Mas, B é a união dos eventos mutuamente exclusivos $(B \cap A)$ e $(B \cap \bar{A})$; logo $P(B) = P(A \cap B) + P(\bar{A} \cap B)$.

Substituindo o valor de $P(\bar{A} \cap B) = P(B) - P(A \cap B)$ na expressão anterior resulta:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Como exercício, o leitor poderá provar o teorema acima para três eventos:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - $$
$$- P(B \cap C) + P(A \cap B \cap C)$$

Sugestão: faça $A \cup B \cup C = (A \cup B) \cup C$ e aplique o teorema da soma.

Note-se que, apesar de se ter postulado a existência do número $P(A)$ e de várias propriedades (teoremas) que esse número possui, nada foi dito quanto a maneira de se calcular $P(A)$. Para esse cálculo, devem ser feitas certas suposições adicionais que conduzem a um método de avaliação da probabilidade, porém, se essas suposições não forem fundamentadas, precisa-se recorrer à experimentação a fim de se encontrar o valor de $P(A)$.

A f (freqüência relativa) será de grande valia para aproximar o cálculo de $P(A)$. Note-se que não se está afirmando que f_A é a mesma coisa que $P(A)$. Contudo, mesmo que a aproximação for muito grosseira, em nada abalará a lógica do modelo estabelecido acima.

1.8 PROBABILIDADES FINITAS DOS ESPAÇOS AMOSTRAIS FINITOS

Seja S um espaço amostral finito $S = \{a_1, a_2, ..., a_n\}$. Considere-se o evento formado por um resultado simples $A = \{a_i\}$.

A cada evento simples $\{a_i\}$ associa-se um número p_i denominado probabilidade de $\{a_i\}$ satisfazendo as seguintes condições:

a) $p_i \geqslant 0 \quad i = 1, 2, ..., n$

b) $p_1 + p_2 + p_3 + ... + p_n = 1$

A probabilidade $P(A)$ de cada evento composto (mais de um elemento) é então definida pela soma das probabilidades dos pontos de A.

Exemplo: Três cavalos, A, B e C, estão em uma corrida; A tem duas vezes mais probabilidade de ganhar que B, e B tem duas vezes mais probabilidade de ganhar que C.

Quais são as probabilidades de vitória de cada um, isto é, $P(A)$, $P(B)$ e $P(C)$?

Faça $P(C) = p$; desde que B tem duas vezes mais probabilidades de ganhar que C, $P(B) = 2p$ e assim $P(A) = 2\ P(B) = 4p$. Como a soma das probabilidades é 1, então:

$$p + 2p + 4p = 1 \quad \text{ou} \quad 7p = 1 \text{ ou } p = \frac{1}{7}\cdot$$

Logo, temos:

$$P(A) = 4p = \frac{4}{7};\ P(B) = 2p = \frac{2}{7} \text{ e } P(C) = p = \frac{1}{7}\cdot$$

Qual seria a probabilidade de B ou C ganhar?

$$P(A \cup B) = P(A) + P(B), \quad \text{axioma III com } A \cap B = \phi.$$

Logo $$P(B \cup C) = P(B) + P(C) = \frac{2}{7} + \frac{1}{7} = \frac{3}{7}\cdot$$

1.9 ESPAÇOS AMOSTRAIS FINITOS EQUIPROVÁVEIS

Quando se associa a cada ponto amostral a mesma probabilidade, o espaço amostral chama-se equiprovável ou uniforme. Em particular, se S contém n pontos, então, a probabilidade de cada ponto será $\frac{1}{n}$.

Por outro lado, se um evento A contém r pontos, então:

$$P(A) = r \cdot \left(\frac{1}{n}\right) = \frac{r}{n}$$

Este método de avaliar $P(A)$ é freqüentemente enunciado da seguinte maneira:

$$P(A) = \frac{\text{n}^{\text{o}} \text{ de vezes em que o evento } A \text{ pode ocorrer}}{\text{n}^{\text{o}} \text{ de vezes em que o Espaço Amostral } S \text{ ocorre}}$$

ou

$$P(A) = \frac{\text{N.C.F (n}^{\text{o}} \text{ de casos favoráveis)}}{\text{N.T.C. (n}^{\text{o}} \text{ total de casos)}}$$

Exemplos:

I) Escolha aleatoriamente (a expressão "aleatória" nos indicará que o espaço é equiprovável) uma carta de um baralho com 52 cartas.

Seja: A = {a carta é de ouros}

B = {a carta é uma figura}

Calcular $P(A)$ e $P(B)$.

$$P(A) = \frac{\text{n}^{\text{o}} \text{ de ouros}}{\text{n}^{\text{o}} \text{ de cartas}} = \frac{13}{52} = \frac{1}{4}$$

$$P(B) = \frac{\text{n}^{\text{o}} \text{ de figuras}}{\text{n}^{\text{o}} \text{ de cartas}} = \frac{12}{52} = \frac{3}{13}$$

Como se observa, o cálculo da probabilidade de um evento reduz-se a um problema de contagem. Assim é que a Análise Combinatória (Teoria de Contagem) tem fundamental importância para se contar o número de casos favoráveis e o total de casos. Na maioria dos problemas tratados neste livro, a combinação é a técnica que pode ser aplicada.

Combinação de r elementos tomados (combinados) p a p ($p \leq r$). Calcula-se por:

$$Cr, p = \binom{r}{n} = \frac{r!}{p!(r-p)!}$$

onde: $r! = r(r-1)(r-2)\ldots 1$

$p! = p(p-1)(p-2)\ldots 1$

admite-se que $0! = 1$

Exemplo: Quantas comissões de três pessoas podem-se formar com um grupo de dez pessoas?

$$C_{10,3} = \binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{3 \cdot 2 \cdot 7!} = 120$$

II) Num lote de 12 peças, 4 são defeituosas; duas peças são retiradas aleatoriamente. Calcule:

a) A probabilidade de ambas serem defeituosas

b) A probabilidade de ambas não serem defeituosas

c) A probabilidade de ao menos uma ser defeituosa.

Solução:

a) A = {ambas são defeituosas}

A pode ocorrer $\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4 \cdot 3 \cdot 2!}{2 \cdot 2!} = 6$ vezes

S pode ocorrer $\binom{12}{2} = \frac{12!}{2!(12-2)!} = \frac{12 \cdot 11 \cdot 10!}{2 \cdot 10!} = 66$ vezes

logo, $P(A) = \frac{\text{N.C.F.}}{\text{N.T.C.}} = \frac{6}{66} = \frac{1}{11}$

b) B = {ambas não defeituosas}

B pode ocorrer $\binom{8}{2} = \frac{8!}{2!(8-2)!} = \frac{8 \cdot 7 \cdot 6!}{2 \cdot 6!} = 28$ vezes

S pode ocorrer $\binom{12}{2} = \frac{12!}{2!(12-2)!} = \frac{12 \cdot 11 \cdot 10!}{2 \cdot 2!} = 66$ vezes

logo, $P(B) = \frac{28}{66} = \frac{14}{33}$

c) C = {ao menos uma é defeituosa}

C é o complemento de B, $C = \bar{B}$

$P(C) = 1 - P(B) = 1 - \frac{14}{33} = \frac{19}{33}$

EXERCÍCIOS – SÉRIE I – CAPÍTULO 1

1. Lance um dado e uma moeda.
   a) Construa o espaço amostral
   b) Enumere os seguintes eventos
      A = {coroa, marcado por número par}
      B = {cara, marcado por número ímpar}
      C = {múltiplos de 3}
   c) Expresse os eventos
      I) $\bar{B}$
      II) A ou B ocorrem
      III) B e C ocorrem
      IV) $\overline{A \cup B}$

d) Verifique dois a dois os eventos A, B e C e diga quais são mutuamente exclusivos.

2. Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{4}$ e A e B mutuamente exclusivos, calcular:

   a) $P(\bar{A})$ b) $P(\bar{B})$ c) $P(A \cap B)$

   d) $P(A \cup B)$ e) $P\overline{(A \cap B)}$

3. Se $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{3}$ e $P(A \cap B) = \frac{1}{4}$
   Calcule

   a) $P(A \cup B)$
   b) $P(\bar{A} \cup \bar{B})$
   c) $P(\bar{A} \cap \bar{B})$

4. Determine a probabilidade de cada evento:

   a) um número par aparece no lançamento de um dado não viciado;
   b) um rei aparece ao extrair-se uma carta de um baralho;
   c) pelo menos uma cara aparece no lançamento de 3 moedas;
   d) pelo menos uma cara aparece no lançamento de "n" moedas;
   e) duas copas aparecem ao retirarem-se duas cartas de um baralho;
   f) uma carta de copas e uma de ouros aparecem ao extraírem-se duas cartas de um baralho.

5. Um número inteiro é escolhido aleatoriamente dentre os números 1, 2, 3, ..., 50. Qual a probabilidade de:

   a) o número ser divisível por 5;
   b) terminar em 3;
   c) ser primo;
   d) ser divisível por 6 ou 8.

6. Qual a probabilidade de sair um rei ou uma carta de copas, quando retiramos uma carta de um baralho?

7. Dois dados são lançados simultaneamente. Qual a probabilidade de:

   a) a soma ser menor que 4;
   b) a soma ser 9;
   c) o primeiro resultado ser maior do que o segundo.

8. Numa urna são misturadas dez bolas numeradas de 1 a 10. Duas bolas são retiradas (*a*, *b*) sem reposição. Qual a probabilidade de $a + b = 10$?

9. Um lote é formado por 10 peças boas, 4 com defeitos e duas com defeitos graves. Uma peça é escolhida ao acaso. Calcule a probabilidade de que:

   a) ela não tenha defeitos graves;

   b) ela não tenha defeitos;

   c) ela ou seja boa ou tenha defeitos graves.

10. Considere o mesmo lote do problema anterior.

    Retiram-se 2 peças ao acaso. Qual a probabilidade de que:

    a) ambas sejam perfeitas;

    b) pelo menos uma seja perfeita;

    c) nenhuma tenha defeito grave;

    d) nenhuma seja perfeita.

11. Uma urna contém 5 bolas brancas e 6 pretas. Três bolas são retiradas. Calcular a probabilidade de:

    a) todas pretas;

    b) exatamente uma branca;

    c) ao menos uma preta.

12. Numa classe existem 5 alunos do 4º ano, 4 do 2º e 3 do 3º ano. Qual a probabilidade de serem sorteados 2 alunos do 2º ano, 3 do 4º e 2 do 3º?

13. Numa urna existem N bolas assim distribuídas: N_V (quantidade de bolas vermelhas); N_A (quantidade de bolas azuis); e N_P (quantidade de bolas pretas). Qual a probabilidade de retirarmos "n" bolas; sendo n_V (nº de bolas vermelhas), n_A (nº de bolas azuis) e n_P (nº de bolas pretas)?

1.10 PROBABILIDADE CONDICIONAL

Seja *E*: lançar um dado, e o evento A = {sair o nº 3}. Então

$$P(A) = \frac{1}{6}$$

Considere agora o evento B = {sair um nº ímpar} = {1, 3, 5}.

É de grande importância para o cálculo das probabilidades se calcular a probabilidade condicional. No exemplo, pode-se estar interessado em avaliar a probabilidade do evento A condicionada à ocorrência do evento B. Em sím-

bolos, designa-se por $P(A/B)$; lê-se: "probabilidade do evento A condicionada à ocorrência B", ou, ainda, "probabilidade de A dado B".

Assim:
$$P(A/B) = \frac{1}{3}$$

Observe: dada a informação da ocorrência de um evento, teremos a redução do espaço-amostra; no exemplo $S = \{1, 2, 3, 4, 5, 6\}$ foi reduzido para $S^* = \{1, 3, 5\}$ e é neste espaço-amostra reduzido que se avalia a probabilidade do evento.

Definição: Dados dois eventos, A e B, denota-se $P(A/B)$ a probabilidade condicionada do evento A, quando B tiver ocorrido, por:

$$\boxed{P(A/B) = \frac{P(A \cap B)}{P(B)}}$$

, com $P(B) \neq 0$, pois B já ocorreu.

Pode-se constatar que $P(A/B)$, assim definida, satisfaz os axiomas da probabilidade já mencionados.

Para aplicações, convém encontrar uma fórmula mais prática para o cálculo da probabilidade condicional; assim:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\dfrac{\text{NCF}(A \cap B)}{\text{NTC}}}{\dfrac{\text{NCF}(B)}{NTC}} = \boxed{\frac{\text{NCF}(A \cap B)}{\text{NCF}(B)}}$$

Dessa maneira, para avaliar a probabilidade de A, dado B, basta contar o número de casos favoráveis ao evento $A \cap B$: [NCF$(A \cap B)$] e dividir esse número pela quantidade de casos favoráveis ao evento B: [NCF(B)].

Exemplo: Dois dados são lançados. Consideremos os eventos:

$A = \{(x_1, x_2)/x_1 + x_2 = 10\}$ e $B = \{(x_1, x_2)/x_1 > x_2\}$

onde x_1 é o resultado do dado 1 e x_2 é o resultado do dado 2.

Avaliar $P(A)$; $P(B)$; $P(A/B)$ e $P(B/A)$.

Solução:

$$S = \begin{Bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{Bmatrix}$$

$$P(A) = \frac{\text{NCF ao evento } A}{\text{NTC}} = \frac{3}{36} = \frac{1}{12} \; // \; P(B) = \frac{\text{NCF a } B}{\text{NTC}} = \frac{15}{36} = \frac{5}{12}$$

$$P(A/B) = \frac{\text{NCF a } (A \cap B)}{\text{NCF a } B} = \frac{1}{15}$$

Note: Apenas o par (6, 4) é favorável ao evento $(A \cap B)$.

$$P(B/A) = \frac{\text{NCF a } (A \cap B)}{\text{NCF a } A} = \frac{1}{3}$$

1.11 TEOREMA DO PRODUTO

A partir da definição de probabilidade condicional pode-se enunciar o teorema do produto:

"A probabilidade da ocorrência simultânea de dois eventos, A e B, do mesmo espaço-amostra, é igual ao produto da probabilidade de um deles pela probabilidade condicional do outro, dado o primeiro."

Assim:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \Rightarrow \quad \boxed{P(A \cap B) = P(B)(A/B)}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \Rightarrow \quad \boxed{P(A \cap B) = P(A)(B/A)}$$

Exemplo: Em um lote de 12 peças, 4 são defeituosas, 2 peças são retiradas uma após a outra sem reposição. Qual a probabilidade de que ambas sejam boas?

Solução: A = {a primeira peça é boa}

B = {a segunda peça é boa}

$$P(A \cap B) = P(A)\,P(B/A) = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33}$$

1.12 INDEPENDÊNCIA ESTATÍSTICA

Um evento A é considerado independente de um outro evento B se a probabilidade de A é igual à probabilidade condicional de A dado B, isto é, se

$$\boxed{P(A) = P(A/B)}$$

É evidente que, se A é independente de B, B é independente de A; assim:

$$P(B) = P(B/A)$$

Considerando o teorema do produto, pode-se afirmar que: se A e B são independentes, então:

$$P(A \cap B) = P(A) \cdot P(B)$$

Dados "n" eventos $A_1, A_2, \ldots, A_n$, diz-se que eles são independentes se o forem 2 a 2; 3 a 3; ...; n a n.

Isto é, se as igualdades abaixo forem verificadas:

$P(A_1 \cap A_2) = P(A_1)\,P(A_2)$; ... ; $P(A_{n-1} \cap A_n) = P(A_{n-1}) \cdot P(A_n)$

$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$; ... ;

$P(A_{n-2} \cap A_{n-1} \cap A_n) = P(A_{n-2}) \cdot P(A_{n-1}) \cdot P(A_n)$

.

.

$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \cdot P(A_3) \ldots P(A_{n-1}) \cdot P(A_n)$

Exemplo 1. Em uma caixa temos 10 peças, das quais 4 são defeituosas. São retiradas duas peças, uma após a outra, com reposição. Calcular a probabilidade de ambas serem boas.

Solução: A = {a primeira peça é boa}

B = {a segunda peça é boa}

Notem: A e B são independentes, pois $P(B) = P(B/A)$.

Logo, $P(A \cap B) = P(A) \cdot P(B) = \frac{6}{10} \cdot \frac{6}{10} = \frac{9}{25}$

Exemplo 2. Sendo $S = \{1, 2, 3, 4\}$ um espaço-amostra equiprovável e $A = \{1, 2\}$; $B = \{1, 3\}$; $C = \{1, 4\}$ três eventos de S. Verificar se os eventos A, B e C são independentes.

Solução: Para A e B $\quad P(A) = \frac{1}{2}$; $P(B) = \frac{1}{2}$; $P(A \cap B) = \frac{1}{4}$;

logo, $P(A \cap B) = P(A) \cdot P(B) = \frac{1}{4}$.

Para A e C $P(A) = \frac{1}{2}$; $P(C) = \frac{1}{2}$; $P(A \cap C) = \frac{1}{4}$;

logo, $P(A \cap C) = P(A) \cdot P(C) = \frac{1}{4}$.

Para B e C $P(B) = \frac{1}{2}$; $P(C) = \frac{1}{2}$; $P(B \cap C) = \frac{1}{4}$;

logo, $P(B \cap C) = P(B) \cdot P(C) = \frac{1}{4}$.

Para A, B e C $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{2}$; $P(C) = \frac{1}{2}$; $P(A \cap B \cap C) = \frac{1}{4}$;

logo, $P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C)$

portanto os eventos A, B e C não são independentes.

1.13 TEOREMA DE BAYES

Seja $A_1, A_2, A_3, \ldots, A_n$, n eventos mutuamente exclusivos tais que $A_1 \cup A_2 \cup A_n = S$. Sejam $P(A_i)$ as probabilidades conhecidas dos vários eventos, e B um evento qualquer de S tal que são conhecidas todas as probabilidades condicionais $P(B/A_i)$.

Então, para cada "i", tem-se:

$$P(A_i/B) = \frac{P(A_i) \cdot P(B/A_i)}{P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + \ldots + P(A_n) \cdot P(B/A_n)}$$

O resultado acima é bastante importante, pois relaciona probabilidades *a priori* $P(A_i)$ com probabilidades *a posteriori* $P(A_i/B)$, probabilidade de A_i depois que ocorrer B.

Exemplo: Admita a seguinte configuração:

Cores \ *Urnas*	u_1	u_2	u_3
PRETAS	3	4	2
BRANCAS	1	3	3
VERMELHAS	5	2	3

Escolheu-se uma urna ao acaso e dela extraiu-se uma bola ao acaso, verificando-se que a bola é branca. Qual a probabilidade da bola ter vindo da urna 2? da 3?

Solução: $P(u_1) = \frac{1}{3}$; $P(u_2) = \frac{1}{3}$; $P(u_3) = \frac{1}{3}$

$P(br/u_1) = \frac{1}{9}$; $P(br/u_2) = \frac{3}{9} = \frac{1}{3}$; $P(br/u_3) = \frac{3}{8}$

deseja-se calcular $P(u_2/\text{br})$ e $P(u_3/\text{br})$.

Aplicando-se o teorema de Bayes, tem-se:

$$P(u_2/br) = \frac{P(u_2) \cdot P(br/u_2)}{P(u_1) \cdot P(br/u_1) + P(u_2) \cdot P(br/u_2) + P(u_3) \cdot P(br/u_3)}$$

$$P(u_2/br) = \frac{\frac{1}{3} \cdot \frac{1}{3}}{\frac{1}{3} \cdot \frac{1}{9} + \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{3}{8}} = \frac{24}{59}$$

ou seja, a probabilidade *a priori* de u_2 era $\frac{1}{3}$. Dada a informação que saiu uma bola branca, a probabilidade *a posteriori* de u_2 será $\frac{24}{59}$.

$$P(u_3/br) = \frac{P(u_3) \cdot P(br/u_3)}{P(u_1) \cdot P(br/u_1) + P(u_2) \cdot P(br/u_2) + P(u_3) \cdot P(br/u_3)} = \frac{27}{59}$$

Como $P(u_1/br) + P(u_2/br) + P(u_3/br) = 1$

tem-se que: $P(u_1/br) = 1 - \left(\frac{24}{59} + \frac{27}{59}\right) = \frac{8}{59}$

EXERCÍCIOS – SÉRIE II – CAPÍTULO 1

1. Dado $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{3}$, $P(A \cap B) = \frac{1}{4}$, calcular:

   a) $P(A \cup B)$; b) $P(A/B)$; c) $P(B/A)$; d) $P[(A \cup B)/B]$

2. Faça A e B serem eventos com $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{3}$ e $P(A \cap B) = \frac{1}{4}$. Encontre $P(\bar{A}/\bar{B})$ e $P(\bar{B}/\bar{A})$.

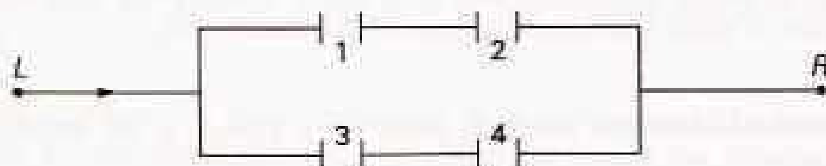
3. Qual a probabilidade de que *r* pessoas façam aniversário em dias distintos?

4. As probabilidades de 3 jogadores marcarem um *penalty* são respectivamente $\frac{2}{3}, \frac{4}{5}$ e $\frac{7}{10}$

Se cada um "cobrar" uma única vez, qual a probabilidade de:

a) todos acertarem;
b) apenas um acertar;
c) todos errarem.

5. A probabilidade de fechamento de cada relé do circuito apresentado abaixo é dada por *p*. Se todos os relés funcionarem independentemente, qual será a probabilidade de que haja corrente entre os terminais *L* e *R*?



6. Uma urna contém 12 bolas: 5 brancas, 4 vermelhas e 3 pretas. Outra contém 18 bolas: 5 brancas, 6 vermelhas e 7 pretas. Uma bola é retirada de cada urna. Qual é a probabilidade de que as duas bolas sejam da mesma cor?

7. Numa bolsa temos 5 moedas de Cr$ 1,00 e 4 de Cr$ 0,50. Qual a probabilidade de, ao retirarmos duas moedas, obtermos Cr$ 1,50?

8. Uma urna contém 5 bolas pretas, três vermelhas e duas brancas. Foram extraídas 3 bolas com reposição. Qual a probabilidade de terem sido duas bolas pretas e uma vermelha?

9. A urna nº 1 contém: 1 bola vermelha e 2 brancas. A urna nº 2 contém: 2 bolas vermelhas e 1 branca. Tiramos aleatoriamente uma bola da urna nº 1, colocamos na urna nº 2 e misturamos. Em seguida tiramos aleatoriamente uma bola da urna nº 2. Qual é a probabilidade de tirarmos uma bola branca da urna nº 2?

10. A urna 1 contém *x* bolas brancas, e *y* bolas vermelhas. A urna 2 contém *z* bolas brancas e *v* vermelhas. Uma bola é escolhida ao acaso da urna 1 e posta na urna 2. A seguir uma bola é escolhida ao acaso da urna 2. Qual será a probabilidade de que esta bola seja branca?

11. Uma urna contém 10 bolas pretas e 5 bolas vermelhas. São feitas retiradas aleatórias. Cada bola retirada é reposta, juntamente com 5 bolas da mesma cor. Qual é a probabilidade de saírem nessa ordem: 1 preta, 1 preta, 1 vermelha, 1 vermelha? E nesta ordem: 1 preta, 1 vermelha, 1 preta, 1 vermelha? Dado que a primeira bola é preta, qual é a probabilidade de que a segunda seja preta?

12. Uma caixa *A* contém 8 peças, das quais 3 são defeituosas e uma caixa *B* contém 5 peças, das quais 2 são defeituosas. Uma peça é retirada aleatoriamente de cada caixa:

    I) Qual a probabilidade p de que ambas as peças não sejam defeituosas?

    II) Qual é a probabilidade p de que uma peça seja defeituosa e a outra não?

    III) Se uma peça é defeituosa e a outra não, qual é a probabilidade p de que a peça defeituosa venha da caixa *A*?

13. A probabilidade de uma mulher estar viva daqui a 30 anos é $\frac{3}{4}$ e de seu marido $\frac{3}{5}$. Calcular a probabilidade de:

    a) apenas o homem estar vivo;
    b) somente a mulher estar viva;
    c) pelo menos um estar vivo;
    d) ambos estarem vivos.

14. Uma urna *A* contém 4 bolas: 2 brancas, 2 pretas; uma urna *B* contém 5 bolas: 3 brancas, 2 pretas. Uma bola é transferida de *A* para *B*. Uma bola é retirada de *B* e verificada ser branca. Qual é a probabilidade de que a bola transferida tenha sido branca?

15. São dadas duas urnas *A* e *B*. A urna *A* contém uma bola preta e uma vermelha. A urna *B* contém duas bolas pretas e três vermelhas. Uma bola é escolhida ao acaso na urna *A* e colocada na urna *B*. Uma bola é então extraída ao acaso, da urna *B*. Pergunta-se:

    a) Qual a probabilidade de que ambas as bolas sejam da mesma cor?
    b) Qual a probabilidade de que a primeira bola seja vermelha, sabendo-se que a segunda foi preta?

16. Uma urna contém 5 bolas vermelhas e 3 brancas. Uma bola é selecionada aleatoriamente da urna e abandonada, e duas de outra cor são colocadas na urna. Uma segunda bola é então selecionada da urna. Encontre a probabilidade de que:

I) a segunda bola seja vermelha; e

II) ambas as bolas sejam da mesma cor.

17. Recorrendo-se ao problema precedente:

I) se a segunda bola é vermelha, qual é a probabilidade de que a primeira bola seja vermelha?

II) se ambas são da mesma cor, qual é a probabilidade de que sejam brancas?

18. A urna *A* contém *x* bolas vermelhas e *y* bolas brancas e a urna *B* contém *z* bolas vermelhas e *v* bolas brancas.

I) Se uma urna é selecionada ao acaso, e uma bola retirada, qual é a probabilidade de que a bola seja vermelha?

II) Se uma bola é retirada da urna *A* e colocada na urna *B*, e uma bola é retirada da urna *B*, qual é a probabilidade de que a segunda bola seja vermelha?

19. Uma urna contém *x* bolas brancas e *y* bolas pretas. Extraem-se todas elas. Qual a probabilidade de que saiam primeiro as brancas e as pretas?

20. Seja *E*: lançar dois dados, e

$A = \{(x_1, x_2)/x_1 + x_2 = 8\}$

$B = \{(x_1, x_2)/x_1 = x_2\}$

$C = \{(x_1, x_2)/x_1 + x_2 = 10\}$

$D = \{(x_1, x_2)/x_1 > x_2\}$

$E = \{(x_1, x_2)/x_1 = 2x_2\}$

Calcular:

a) $P(A/B)$	e) $P(C/E)$	i) $P(A/E)$
b) $P(C/D)$	f) $P(C/A)$	j) $P(B/E)$
c) $P(D/E)$	g) $P(A/D)$	k) $P(A/B \cap C)$
d) $P(A/C)$	h) $P(B/C)$	l) $P(A \cap B/C \cap D)$

21. Temos duas caixas: na primeira há 3 bolas brancas e 7 pretas e na 2ª, 1 bola branca e 5 pretas. De uma caixa escolhida ao acaso, seleciona-se uma bola e verifica-se que é preta. Qual a probabilidade de que a caixa de onde for extraída a bola seja a primeira? e a segunda?

22. A probabilidade de um indivíduo de classe *A* comprar um carro é $\frac{3}{4}$, de *B* é $\frac{1}{6}$ e de *C* é $\frac{1}{20}$. A probabilidade do indivíduo de classe *A* comprar

um carro da marca D é $\frac{1}{10}$; de B comprar da marca D é $\frac{3}{5}$ e de C é $\frac{3}{10}$. Em certa loja comprou-se um carro da marca D. Qual a probabilidade de que o indivíduo da classe B o tenha comprado?

23. Em certo colégio, 5% dos homens e 2% das mulheres têm mais do que 1,80 m de altura. Por outro lado, 60% dos estudantes são homens. Se um estudante é selecionado aleatoriamente e tem mais de 1,80 m de altura, qual a probabilidade de que o estudante seja mulher?

24. Três máquinas, A, B e C produzem respectivamente 40%, 50% e 10% do total de peças de uma fábrica. As porcentagens de peças defeituosas nas respectivas máquinas são 3%, 5% e 2%. Uma peça é sorteada ao acaso e verifica-se que é defeituosa. Qual a probabilidade de que a peça tenha vindo da máquina B?

25. Apenas uma em cada dez pessoas de uma população tem tuberculose. Das pessoas que têm tuberculose 80% reagem positivamente ao teste Y, enquanto apenas 30% dos que não têm tuberculose reagem positivamente. Uma pessoa da população é selecionada ao acaso e o teste Y é aplicado. Qual a probabilidade de que essa pessoa tenha tuberculose, se reagiu positivamente ao teste?

EXERCÍCIOS – SÉRIE III – CAPÍTULO 1

1. Uma moeda é lançada três vezes. Ache a probabilidade de se obterem:
   a) três caras;
   b) duas caras e uma coroa;
   c) uma cara;
   d) pelo menos uma coroa;
   e) nenhuma cara.

2. São lançados dois dados. Qual a probabilidade de:
   a) obter-se um par de pontos iguais;
   b) um par de pontos diferentes;
   c) um par em que o 1º < 2º;
   d) a soma dos pontos ser um número par;
   e) obter-se soma 7, se o par de pontos é diferente;
   f) obter-se soma 6, dado que o par de pontos é igual;
   g) a soma ser 14.

3. A probabilidade de o aluno *X* resolver esse problema é $\frac{3}{5}$ e a do aluno *Y* é $\frac{4}{7}$. Qual probabilidade de que o problema seja resolvido?

4. No lançamento de um dado, qual a probabilidade de sair o número 5 ou um número par?

5. Um grupo de 15 elementos apresenta a seguinte composição:

	Homens	***Mulheres***
Menores	5	3
Adultos	5	2

Um elemento é escolhido ao acaso. Pergunta-se:

a) Qual a probabilidade de ser homem?
b) Qual a probabilidade de ser adulto?
c) Qual a probabilidade de ser menor e mulher?
d) Sabendo-se que o elemento escolhido é adulto, qual a probabilidade de ser homem?
e) Dado que a escolhida é mulher, qual a probabilidade de ser menor?

6. Um número é escolhido ao acaso no conjunto {1, 2, 3, ..., 20}. Verificar se são independentes os eventos:

a) *X*: o número é múltiplo de 3.
   Y: o número é par.
b) *M*: o número é primo.
   N: o número é ímpar.

7. Um grupo de 100 pessoas apresenta, de acordo com o sexo e filiação partidária, a seguinte composição:

	Partido X	***Partido Y***
Homens	21	39
Mulheres	14	26

Calcular:

a) a probabilidade de um escolhido ser homem;
b) a probabilidade de um escolhido ser mulher do partido Y;
c) a porcentagem dos partidários do Y;

d) a porcentagem dos homens filiados à X;

e) se o sorteado for da X, qual a probabilidade de ser mulher;

f) se o sorteado for homem, qual a probabilidade de ser do Y.

8. Prove: Se A e B são eventos independentes e mutuamente exclusivos, então $P(A) = 0$ ou $P(B) = 0$.

9. Prove: Se A e B são eventos independentes e de probabilidades não nulas, então A e B não são mutuamente exclusivos, isto é, $A \cap B \neq \phi$.

10. Prove: Os eventos A e S são independentes.

11. Prove: Os eventos A e ϕ são independentes.

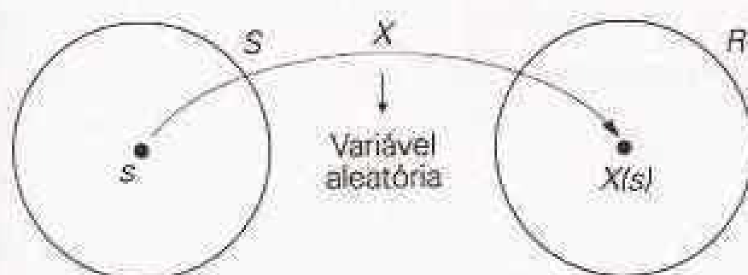
12. Prove: Os eventos S e ϕ são independentes.

2

Variável Aleatória

2.1 DEFINIÇÃO

Sejam *E* um experimento e *S* o espaço associado ao experimento. Uma função *X*, que associe a cada elemento $s \in S$ um número real $X(s)$ é denominada variável aleatória. Veja a ilustração.



Exemplo: *E*: lançamento de duas moedas;

X: nº de caras obtidas nas duas moedas;

$S = \{(c, c), (c, k), (k, c), (k, k)\}$

$X = 0 \rightarrow$ corresponde ao evento (c, c) com probabilidade $\frac{1}{4}$

$X = 1 \rightarrow$ corresponde ao evento (k, c), (c, k) com probabilidade $\frac{2}{4}$

$X = 2 \rightarrow$ corresponde ao evento (k, k) com probabilidade $\frac{1}{4}$

Observações importantes:

1. Observe que, apesar da infelicidade da terminologia "variável aleatória", é uma função cujo domínio é S e contradomínio R.
2. Nas aplicações, é conveniente trabalhar com números e não com eventos, daí, o uso da variável aleatória.
3. Se S é numérico, então $X(s) = s$.
4. Uma *variável aleatória X será discreta* se o número de valores possíveis de X (seu contradomínio) for finito ou infinito numerável. Caso seu contradomínio seja um intervalo ou uma coleção de intervalos, *ela será uma variável contínua.*

2.2 FUNÇÃO DE PROBABILIDADES

Seja X uma variável aleatória discreta.

A probabilidade de que a variável aleatória X assuma um particular valor x, é a função de probabilidade de X que se representa por $P(X = x)$ ou simplesmente $P(x)$. ($\Sigma P(x_i) = 1$). A função $P(X = x)$ determina a distribuição de probabilidades da variável aleatória.

$P(x)$ pode ser expressa por uma tabela, gráfico ou fórmula.

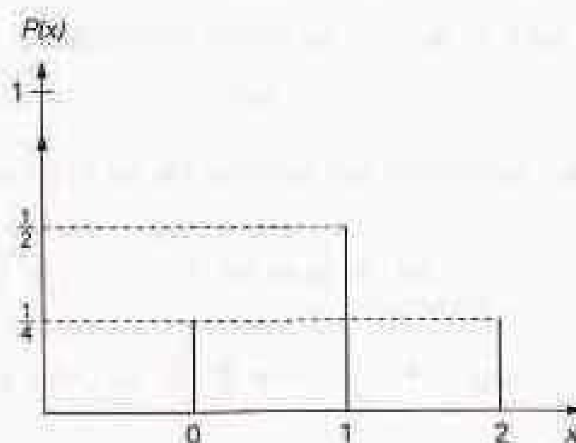
Exemplo: E: lançamento de duas moedas.
X: nº de caras obtidas.

Eis as várias expressões para $P(x)$:

1. *Tabela*

x	0	1	2
$P(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

2. *Gráfico*



3. *Fórmula*

$$P(x) = \frac{1}{4}\binom{2}{x}, \quad \text{para} \quad x = 0, 1, 2.$$

Qualquer função de uma variável aleatória (V.A.) é também uma variável aleatória, isto é, se X é V.A., então $Y = \varphi(x)$ também será.

Exemplo: $X \rightarrow$ V.A. pontos de um dado

$Y = X + X \rightarrow$ V.A. soma dos pontos de dois lançamentos

$Z = \text{Max}\ \{(x_1, x_2)\}$ onde (x_1, x_2) pontos de dois dados.

A distribuição de probabilidade de X dada por uma tabela será:

x	1	2	3	4	5	6
$P(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Para se obter a distribuição de probabilidade de Y convém construir o espaço amostral para y:

$$S = \begin{Bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{Bmatrix}$$

Logo, a distribuição de probabilidade de y será:

y	2	3	4	5	6	7	8	9	10	11	12
$P(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Enquanto a distribuição de Z poderá ser expressa pela tabela:

z	1	2	3	4	5	6
$P(z)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

2.3 FUNÇÃO DE REPARTIÇÃO

Seja X uma variável aleatória discreta.

Define-se Função de Repartição da variável aleatória X, no ponto x, como sendo a probabilidade de que X assuma um valor menor ou igual a x, isto é:

$$\boxed{F(x) = P(X \leq x)}$$

Propriedades:

1. $F(x) = \sum_{x_i \leq x} P(x_i)$ (cálculo de $F(x)$)
2. $F(-\infty) = 0$
3. $F(+\infty) = 1$
4. $P(a < X \leq b) = F(b) - F(a)$
5. $P(a \leq X \leq b) = F(b) - F(a) + P(X = a)$
6. $P(a < X < b) = F(b) - F(a) - P(X = b)$
7. $F(x)$ é contínua à direita $\rightarrow \lim_{x \to x_0} F(x) = F(x_0)$
8. $F(x)$ é descontínua à esquerda, nos pontos em que a probabilidade é diferente de zero.

   $\lim_{x \to x_0} F(x) \neq F(x_0)$ para $P(X = x_0) \neq 0$
9. A função é não decrescente, isto é, $F(b) \geq F(a)$, para $b > a$.

Exemplo: Admita que a variável aleatória X tome os valores 0, 1, 2 com probabilidades $\frac{1}{3}, \frac{1}{6}, \frac{1}{2}$ respectivamente.

Então,

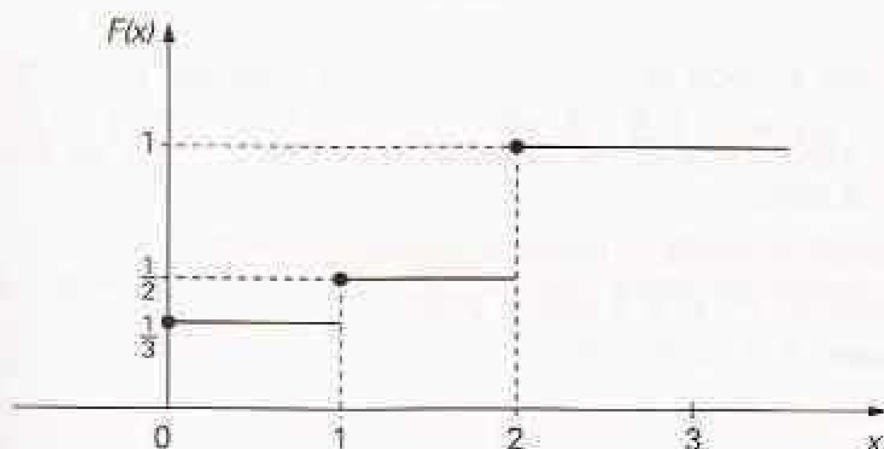
$F(x) = 0$ se $x < 0$

$F(x) = \frac{1}{3}$ se $0 \leq x < 1$

$F(x) = \frac{1}{2}$ se $1 \leq x < 2$

$F(x) = 1$ se $x \geq 2$

Eis o gráfico de $F(x)$



EXERCÍCIOS – SÉRIE I – CAPÍTULO 2

1. No lançamento simultâneo de dois dados, considere as seguintes variáveis aleatórias:

   X = número de pontos obtidos no primeiro dado.

   Y = número de pontos obtidos no segundo dado.

   a) Construir a distribuição de probabilidade através de uma tabela e gráfico das seguintes variáveis

      I) $W = X - Y$

      II) $A = 2Y$

      III) $Z = X\,Y$

      IV) B = máximo de (X, Y)

   b) Construir a Função Repartição das variáveis W, Z e B e fazer os respectivos gráficos.

   c) Aplicando as Propriedades da Função Repartição, calcular as seguintes probabilidades:

      I) $P(-3 < W \leq 3)$

      II) $P(0 \leq W \leq 4{,}5)$

      III) $P(A > 6)$

      IV) $P(Z \leq 5{,}5)$

      V) $P(Z = 3)$

      VI) $P(1 \leq B \leq 4)$

      VII) $P(W \leq -8)$

      VIII) $P(A \geq 11)$

      IX) $P(20 \leq Z \leq 35)$

      X) $P(B = 8)$

      XI) $P(-1 < A < 8)$

      XII) $P(3{,}5 < Z < 34)$

2. Uma variável aleatória discreta tem a distribuição de probabilidade dada por:

$$P(X) = \frac{K}{x} \qquad \text{para} \qquad x = 1, 3, 5, 7$$

a) Calcular o valor de K b) Calcular $P(x = 5)$

3. Seja Z a variável aleatória correspondente ao número de pontos de uma peça de dominó.
   a) Construir a tabela e traçar o gráfico de $P(Z)$.
   b) Determinar $F(Z)$ e traçar o gráfico.
   c) Calcular $P(2 \leq Z < 6)$.
   d) Calcular $F(8)$.

4. Numa sala temos cinco rapazes e quatro moças. São retiradas, aleatoriamente, três pessoas. Faça X a Variável Aleatória número de rapazes.
   a) Determine a distribuição de probabilidade da variável X. Construa uma tabela.
   b) Determine a função repartição de X.
   c) Construa o gráfico de F (X).
   d) Calcule as probabilidades:
      I) $P(X \leq 2)$
      II) $P(X \leq 0)$
      III $P(1 < X \leq 3)$
      IV) $P(2 < X < 3)$
      V) $P(X > 2)$
      VI) $P(X > -1)$
      VII) $P(X < 5)$
   e) Determine: $F(2{,}5)$; $F(3)$; $F(0{,}5)$; $F(3{,}5)$; $F(2)$; $F(1)$; $F(6)$; $F(-0{,}5)$.

2.4 VARIÁVEL ALEATÓRIA CONTÍNUA

Suponha que X é uma variável aleatória, cujo contradomínio é um contínuo de números, um intervalo, por exemplo $0 \leq x \leq 5$, ou uma coleção de intervalos. Neste caso, a variável aleatória é contínua.

Relembrando, no caso da variável aleatória discreta definiu-se $P(x)$ (função probabilidade), como uma função que associa a cada elemento um número não negativo, $P(x_i) = P(X = x_i)$, $i = 1, 2, \ldots$, cuja soma é igual a 1. Se for utilizado o mesmo conceito para a variável aleatória contínua, não se pode indagar qual a probabilidade do i-ésimo valor de X, pois os valores possíveis

de X não são numeráveis, daí $P(x_i)$ não ter sentido. Há necessidade de definir outro conceito, o que se passa a fazer.

2.5 FUNÇÃO DENSIDADE DE PROBABILIDADE

Seja X uma variável aleatória contínua. A função densidade de probabilidade $f(x)$ é uma função que satisfaz as seguintes condições:

a) $f(x) \geqslant 0$ para todo $x \in Rx$.

b) $\int_{Rx} f(x)\, dx = 1$

Além disso, define-se, para qualquer $a < b$ em Rx

$$P(a < X < b) = \int_a^b f(x)\, dx$$

em que Rx é o contradomínio de X.

Observações importantes:

1. A definição acima mostra que a probabilidade de qualquer valor especificado de X, por exemplo x_0, tem $P(X = x_0) = 0$, pois

$$P(X = x_0) = \int_{x_0}^{x_0} f(x)\, dx = 0$$

Sendo assim, as probabilidades abaixo serão todas iguais, se X for variável aleatória contínua:

$$P(a \leqslant X \leqslant b) = P(a \leqslant X < b) = P(a < X \leqslant b) = P(a < X < b)$$

2. Note-se que $f(x)$, densidade de probabilidade, não é probabilidade. Somente quando a função for integrada entre dois limites, ela produzirá uma probabilidade, que será a área sob a curva da função entre $x = a$ e $x = b$; $a < b$

3. Pode-se fazer uma analogia com a Mecânica e considerar-se que numa variável aleatória discreta, a massa de probabilidade está concentrada em pontos isolados da reta real, e, no caso da variável aleatória contínua, a massa de probabilidade está espalhada de modo contínuo em segmentos da reta real;

4. Quanto à função Repartição, neste caso ela é definida como:

$$F(x) = \int_{-\infty}^{x} f(x)\,dx$$

Pode-se provar que $f(x) = \dfrac{dF(x)}{dx}$ para todo x no qual F seja derivável.

Como $F(+\infty) = 1$, devemos ter sempre $\int_{-\infty}^{+\infty} f(x)\,dx = 1$, isto é, a área total debaixo da curva de probabilidade vale sempre 1.

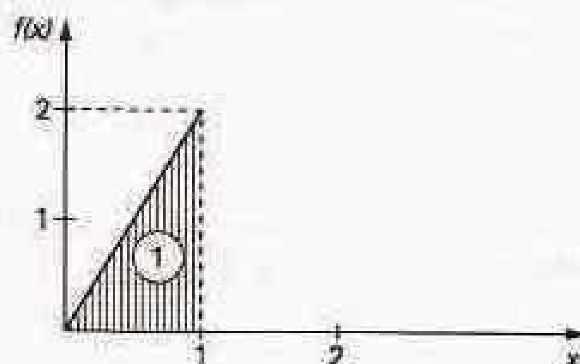
Exemplo 1. Seja X uma variável aleatória contínua. Com a seguinte função densidade de probabilidade:

$$f(x) = \begin{cases} 2x \text{ para } 0 < x < 1 \\ 0 \text{ para quaisquer outros valores} \end{cases}$$

a) Como se vê, $f(x)$ assim definida, é realmente uma função densidade, pois

$f(x) \geq 0$ e para todo $x \in Rx$ $\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{0} dx + \int_{0}^{1} 2x\,dx + \int_{0}^{\infty} dx = 1$

Seu gráfico será



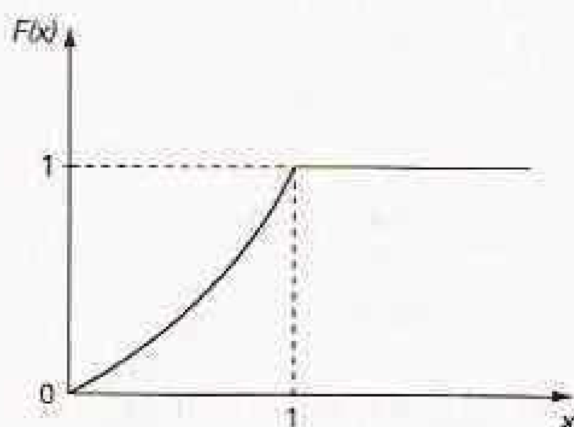
Quanto a $F(x)$ tem-se:

para $x < 0$ $\quad F(x) = \int_{-\infty}^{x} 0\,dx = 0$

para $0 \leq x < 1$ $\quad F(x) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{x} 2x\,dx = x^2$

para $x \geq 1$ $\quad F(x) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} 2x\,dx + \int_{1}^{\infty} 0\,dx = 1$

Cujo gráfico será:



O gráfico de $F(x)$ no caso de V.A. discreta é constituído por segmentos de retas horizontais (degraus), e no caso da V.A. contínua, ele é contínuo para todo x.

b) Qual será o resultado de $P\left(\frac{1}{4} < x < \frac{3}{4}\right)$?

$$P\left(\frac{1}{4} < x < \frac{3}{4}\right) = \int_{\frac{1}{4}}^{\frac{3}{4}} 2\,x dx = \frac{2x^2}{2} \Bigg|_{\frac{3}{4}}^{\frac{1}{4}} = \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = \frac{1}{2}$$

Exemplo 2 Uma variável aleatória tem a seguinte função densidade de probabilidade:

$x < 0 \qquad f(x) = 0$

$0 \leq x < 1 \qquad f(x) = kx^2$

$x \geq 1 \qquad f(x) = 0$

Pede-se determinar k e a função repartição.

Sabe-se que:

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

Então:

$$\int_{-\infty}^{0} 0\, dx + \int_{0}^{1} kx^2 dx + \int_{1}^{\infty} 0\, dx = 1, \text{ ou seja, } \frac{kx^3}{3}\Bigg|_{0}^{1} = \frac{k}{3} = 1 \quad \therefore \quad k = 3$$

Quanto à função repartição, tem-se:

para $x < 0$, $F(x) = \int_{-\infty}^{x} 0 \, dx = 0$

para $0 \leq x < 1$

$$F(x) = \int_{-\infty}^{0} 0 \, dx + \int_{0}^{x} 3x^2 dx = x^3 \Big|_0^x = x^3$$

para $x \geq 1$

$$F(x) = \int_{-\infty}^{0} 0 dx + \int_{0}^{1} 3x^2 \, dx + \int_{1}^{x} 0 \, dx = 0 + 1 + 0 = 1$$

EXERCÍCIOS – SÉRIE II – CAPÍTULO 2

1. Seja $f(x) = \begin{cases} \frac{3}{2}(1 - x^2), \ 0 < x < 1 \\ 0, \text{ caso contrário} \end{cases}$

   Ache a função Repartição e esboce o gráfico.

2. Seja $f(x) = \begin{cases} \frac{1}{2}x \text{ se } 0 \leq x \leq 2 \\ 0 \text{ se caso contrário} \end{cases}$

   Ache a função Repartição e esboce o gráfico.

3. Seja $f(x) = \begin{cases} \frac{1}{6}x + K \text{ se } 0 \leq x \leq 3 \\ 0 \quad \text{em qualquer outro caso} \end{cases}$

   Pede-se:

   a) encontre K b) encontre $P(1 \leq x \leq 2)$

4. Uma variável aleatória contínua X tem a seguinte função densidade de probabilidade:

$x < 0$	$f(x) = 0$
$0 \leq x < 2$	$f(x) = K$
$2 \leq x < 4$	$f(x) = K(x - 1)$
$x \geq 4$	$f(x) = 0$

   Pede-se:

   a) qual o valor de K? b) encontre $F(x)$ e faça o gráfico

5. A função densidade de probabilidade de uma variável aleatória contínua X é proporcional a $x(1-x)$ para $0 < x < 1$, e é zero para outros valores de x. Pede-se:

a) mostre que $f(x) = 6x(1-x)$ para $0 < x < 1$

b) calcule a função repartição

c) calcule $P\left(x \leq \dfrac{1}{2}\right)$

6. Seja x uma variável aleatória contínua tal que:

$f(x) = 0$	para	$x < 0$
$f(x) = Ax$	para	$0 \leq x < 500$
$f(x) = A(100 - x)$	para	$500 \leq x < 1000$
$f(x) = 0$	para	$x \geq 1000$

Determinar

a) o valor da constante A;

b) $P(250 \leq x \leq 750)$

7. Dada a função de distribuição (repartição)

$F(x) = 0$	para	$x < -1$
$F(x) = \dfrac{x+1}{2}$	para	$-1 \leq x < 1$
$F(x) = 1$	para	$1 \leq x$

Calcule:

a) $P\left(-\dfrac{1}{2} < x \leq \dfrac{1}{2}\right)$ b) $P(x = 0)$ c) $P(2 < x \leq 3)$

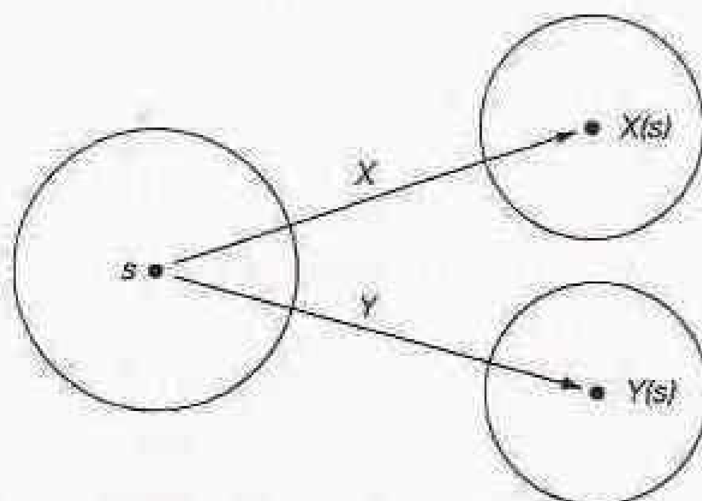
2.5 VARIÁVEL ALEATÓRIA BIDIMENSIONAL

Até aqui considerou-se que o resultado do experimento seria registrado como um único número x. Todavia, existem casos em que há interesse por dois resultados simultâneos. Por exemplo, estatura H e peso P de pessoas. Para tanto precisa-se da seguinte definição:

Sejam: E um experimento aleatório e S o espaço amostral associado a E.

Sejam: $X = X(s)$ e $Y = Y(s)$, duas funções, cada uma associando um número real a cada resultado $s \in S$, *denomina-se* (X, Y) uma variável aleatória bidimensional.

O gráfico ilustra a definição.



Tal como a variável aleatória unidimensional, (X, Y) poderá ser discreta ou contínua, valendo as mesmas considerações feitas anteriormente.

2.7 DISTRIBUIÇÃO CONJUNTA DE DUAS VARIÁVEIS ALEATÓRIAS

Seja (X, Y) uma variável aleatória bidimensional discreta.

Função de Probabilidade: Trata-se de uma função que associa um número $p(x_i, y_j)$ representado por $P(X = x_i, Y = Y_j)$ satisfazendo as condições:

1. $p(x_i, y_j) \geq 0$

2. $\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} p(x_i, y_j) = 1$

Como no caso da variável unidimensional, a distribuição poderá ser representada por uma tabela, gráfico ou fórmula.

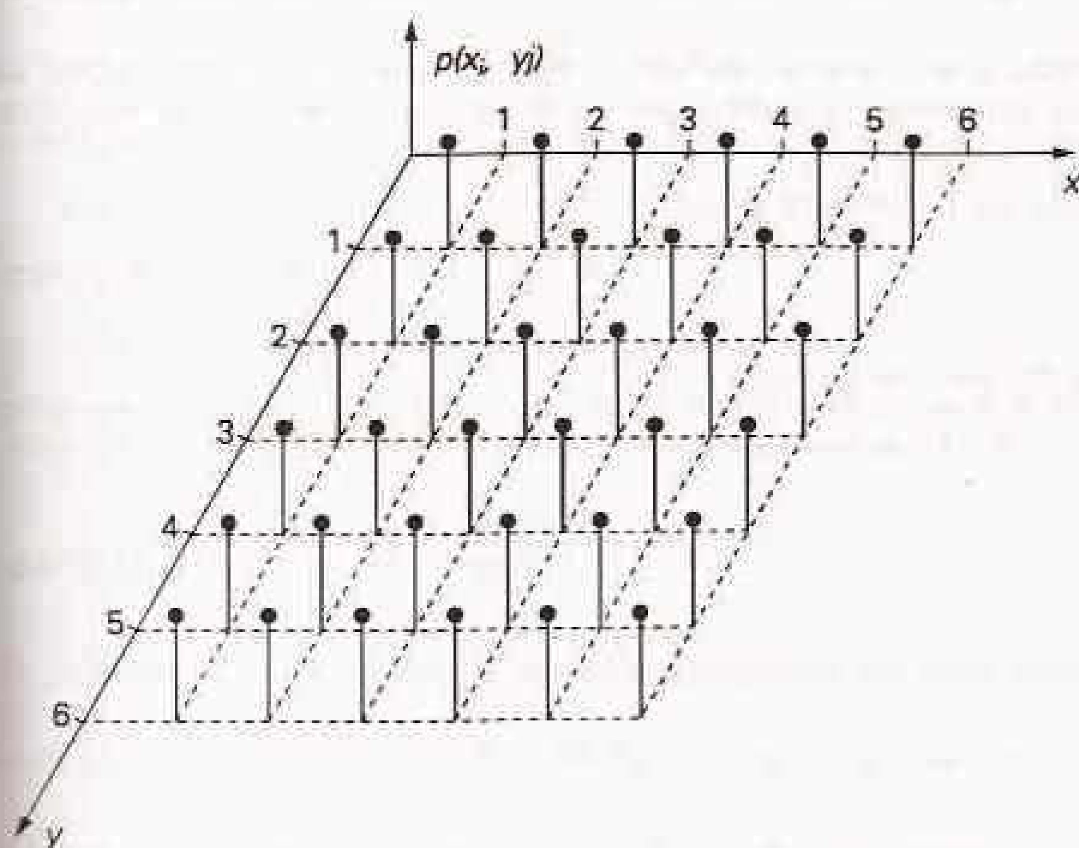
Exemplo: E = jogar dois dados

(X, Y) = pontos dos respectivos dados

1. *Tabela*

x \ y	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

2. *Gráfico*



3. *Fórmula*

$$P(x_i, y_j) = \frac{1}{36} \qquad \begin{array}{l} i = 1, 2, \ldots, 6 \\ j = 1, 2, \ldots, 6 \end{array}$$

2.8 FUNÇÃO DENSIDADE DE PROBABILIDADE CONJUNTA

Seja (x, y) uma variável aleatória bidimensional contínua. Diz-se que $f(x, y)$ é uma função densidade de probabilidade conjunta se:

1. $f(x, y) \geq 0$

2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$

2.9 FUNÇÃO DE REPARTIÇÃO CONJUNTA

É definida como na variável aleatória unidimensional, assim:

$$F(x, y) = P(X \leq x, Y \leq y)$$

Caso (X, Y) discreta, tem-se $F(x, y) = \Sigma\Sigma p(x_i, y_j)$, tal que $x_i \leq x$ e $y_j \leq y$.

Caso (X, Y) contínua, tem-se $F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y)\, dx\, dy$

2.10 DISTRIBUIÇÃO DE PROBABILIDADE MARGINAL

Dada uma variável aleatória bidimensional e sua distribuição conjunta pode-se determinar a distribuição de X sem considerar Y, ou vice-versa. São as chamadas distribuições marginais.

Seja (X, Y) discreta; então:

Distribuição marginal de X: $\begin{cases} P(X = x_i) = P(X = x_i, \ -\infty < y < +\infty) \\ \text{ou} \\ P(X = x_i) = \sum_j P(x_i, y_j) \end{cases}$

Distribuição marginal de Y: $\begin{cases} P(Y = y_j) = P(-\infty < x < \infty, \ Y = y_j) \\ \text{ou} \\ P(Y = y_j) = \sum_i P(x_i, y_j) \end{cases}$

Se a distribuição de (X, Y) é dada por uma tabela, as probabilidades marginais serão dadas pela soma das linhas ou colunas.

No caso de (X, Y) contínua, teremos:

$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \rightarrow$ função densidade marginal de X

$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \rightarrow$ função densidade marginal de Y

2.11 VARIÁVEIS ALEATÓRIAS INDEPENDENTES

Seja (X, Y) uma variável aleatória discreta bidimensional. Diz-se que X e Y são independentes se, e somente se, $p(x_i, y_j) = p(x_i)\, p(Y_j)$ para quaisquer i e j.

Seja (X, Y) uma variável aleatória contínua bidimensional. Diz-se que X e Y são independentes se, e somente se, $f(x, y) = g(x)\, h(y)$ para todo (x, y).

Exemplo: dada a distribuição de probabilidade conjunta de (X, Y) pela tabela abaixo. Verificar se X e Y são independentes.

$y \backslash x$	0	1	2	$p(y_j)$
0	0,10	0,20	0,20	0,50
1	0,04	0,08	0,08	0,20
2	0,06	0,12	0,12	0,30
$p(x_i)$	0,20	0,40	0,40	1,00

Solução: Para todo i, j; $i = 0, 1, 2$ e $j = 1, 2, 3$ deve-se ter:

$$p(x_i, y_j) = p(x_i)\, p(y_j)$$

Para o par (0,0) tem-se:

$$p(0{,}0) = 0{,}10 \text{ e } p(x = 0) \cdot p(y = 0) = 0{,}2 \cdot 05 = 0{,}10$$

Para o par (2,1) tem-se:

$$p(2{,}1) = 0{,}08 \text{ e } p(x = 2) \cdot p(y = 1) = 0{,}4 \cdot 0{,}2 = 0{,}08$$

Continuando-se os cálculos para todos os pares verifica-se que as igualdades são constatadas, daí concluir-se que x e y são independentes.

2.12 MEDIDAS DE POSIÇÃO

2.12.1 Média ou esperança matemática

Define-se Esperança Matemática ou Média de uma variável aleatória discreta como:

$$E[X] = \mu_x = \mu = \sum_i x_i\, P(x_i)$$

Para o caso de variável aleatória contínua tem-se:

$$E[X] = \mu_X = \mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

Exemplo: E = lançamento de um dado

X = ponto obtido

X = 1, 2, 3, 4, 5, 6

$$P(X) = \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$$

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3{,}5$$

Propriedades da Média

Serão demonstradas para o caso de variáveis discretas.

1. A média de uma constante é a própria constante.

$$E[K] = \sum_i KP(x_i) = K \sum_i P(x_i) = K$$

2. Multiplicando uma variável aleatória X por uma constante, sua média fica multiplicada por essa constante.

$$E[KX] = \sum_i Kx_i P(x_i) = K \sum_i x_i P(x_i) = KE[K]$$

3. A média da soma ou da diferença de duas variáveis aleatórias é a soma ou diferença das médias.

$$E[x \pm y] = \sum_i \sum_j (x_i \pm y_j)\, P(x_i, y_j) = \sum_i \sum_j x_i\, P(x_i, y_j) \pm$$

$$\pm \sum_i \sum_j y_j\, P(x_i, y_j) =$$

$$= \sum_i x_i \sum_j P(x_i, y_j) \pm \sum_j y_j \sum_i P(x_i, y_j) = \sum_i x_i\, P(x_i) \pm$$

$$\pm \sum_j y_j\, P(y_j) = E[X] \pm E[Y].$$

4. Somando ou subtraindo uma constante a uma variável aleatória, a sua média fica somada ou subtraída da mesma constante.

$$E[X \pm K] = E[X] \pm E[K] = E[X] \pm K$$

5. A média de uma variável aleatória centrada é zero.

Diz-se que a variável aleatória está centrada quando se calculam todos os desvios $(x_i - \mu_X)$. Assim:

$$E[X - \mu_{(X)}] = E[X] - E[\mu_{(X)}] = \mu_X - \mu_X = 0$$

6. A média do produto de duas variáveis aleatórias independentes é o produto das médias.

$$E[XY] = \sum_i \sum_j X_i \cdot Y_j \cdot P(X_i, Y_j)$$

$$E[XY] = \sum_i \sum_j X_i \cdot Y_j \cdot P(X_i) \cdot P(Y_j), \text{ pois } X \text{ e } Y \text{ são independentes.}$$

$$E[XY] = \sum_i X_i P(X_i) \sum_j Y_j P(Y_j)$$

$$E[XY] = E[X] \cdot E[Y]$$

2.12.2 Mediana

Mediana de uma variável aleatória é o valor que divide a distribuição em duas partes iguais, ou seja, $F(Md) = 0{,}5$, onde Md = mediana.

Exemplo: Seja X uma variável aleatória com a seguinte função repartição:

$F(x) = 0$	para	$x < 0$
$F(x) = x^2$	para	$0 \leq x < 1$
$F(x) = 1$	para	$x \geq 1$

Logo a mediana será o valor de x tal que $F(x = Md) = \frac{1}{2}$

Assim: $Md^2 = \frac{1}{2}$ $\qquad Md = \sqrt{\frac{1}{2}}$

2.12.3 Moda

É o valor da variável com maior probabilidade, se X for discreta, ou maior densidade se X for contínua.

Exemplo:

a) X é uma variável aleatória discreta, tal que:

X	-1	0	2
$P(X)$	0,3	0,2	0,5

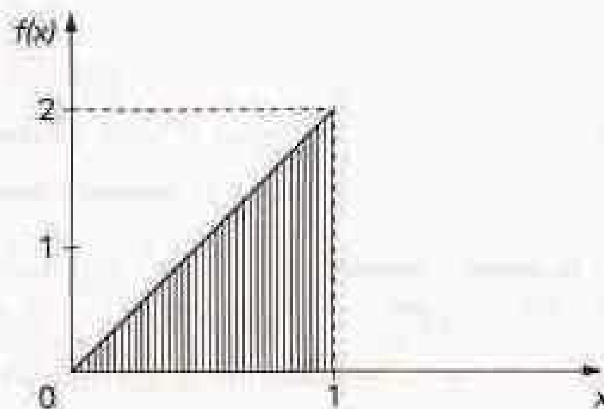
Logo, a moda será o 2.

b) X é uma variável aleatória contínua, tal que:

$f(x) = 2x$ para $0 \leq x \leq 1$

$f(x) = 0$ para outros valores

O gráfico de $f(x)$ é:



Então: *Moda* $m_0 = 1$

(observe o gráfico)

Mediana:

$F(Md) = 0,5$

logo

$$\int_0^{Md} 2x\, dx = 0,5 \Rightarrow \left.\frac{2x^2}{2}\right|_0^{Md} \quad Md^2 = 0,5$$

portanto $Md = \sqrt{\frac{1}{2}}$

2.13 MEDIDAS DE DISPERSÃO

2.13.1 Variância

Define-se variância de uma variável aleatória como sendo:

$$Var[X] = \sigma^2_{(X)} = E[(X - \mu_{(X)})^2]$$

Para X discreta: $\sigma^2_{(X)} = \sum (X_i - \mu_{(X)})^2 \, P(x_i)$

Para X contínua: $\sigma^2_{(X)} = \int_{-\infty}^{\infty} (x_i - \mu_{(X)})^2 \; f(x) \; dx$

2.13.2 Desvio-padrão

Por definição, desvio-padrão é a raiz quadrada da variância:

$$\sigma_{(X)} = \sqrt{\sigma^2_{(X)}}$$

Pode-se encontrar uma fórmula mais prática para o cálculo da variância:

$$\sigma^2_{(X)} = E[(X - \mu_{(X)})^2] = E[(X^2 - 2X\mu_X + \mu_X^2)] =$$

$$= E[X^2] - E[2X\mu_X] + E[\mu_X^2] =$$

$$= E[X^2] - 2\mu_X E[X] + \mu_X^2 =$$

$$= E[X^2] - 2\mu_X \cdot \mu_X + \mu_X^2$$

$$\therefore \; \sigma^2_{(X)} = E[X^2] - \mu^2_{(X)}$$

Propriedades da Variância

Serão demonstradas para o caso de variáveis discretas.

1. A variância de uma constante é zero.

$\sigma^2_{(K)} = E[(K - \mu_{(K)})^2] = E[(K - K)^2] = 0$

2. Multiplicando-se uma variável aleatória por uma constante, sua variância fica multiplicada pelo quadrado da constante.

$$\sigma^2_{(kx)} = E[(KX - \mu_{(kx)})^2] = E[K^2(X - \mu_x)^2] =$$

$$= K^2E[(X - \mu_x)^2] = K^2\,\sigma^2_{(x)}$$

3. Somando-se ou subtraindo-se uma constante à uma variável aleatória, sua variância não se altera.

$$\sigma^2\ (X \pm K) = \sigma^2(X) + \sigma^2(K) = \sigma^2(X) \qquad \text{pois } \sigma^2(K) = 0$$

4. A variância da soma ou diferença de duas variáveis aleatórias independentes é a soma das respectivas variâncias.

$$\sigma^2\,(X \pm Y) = E[[(X \pm Y) - (\mu_x \pm \mu_y)]^2] =$$

$$= E[[(X - \mu_x) \pm (Y - \mu_y)]^2] =$$

$$= E[(X - \mu_x)^2] \pm 2 \cdot E[(X - \mu_x) \cdot (Y - \mu_y)] + E[(Y - \mu_y)^2]$$

mas, $E[(X - \mu_x) \cdot (Y - \mu_y)] = E[(X - \mu_x)] \cdot E[(Y - \mu_y)] =$

$= COV_{xy} = 0$, pois X e Y são independentes. Onde COV_{xy} = covariância entre x e y. Veja na próxima seção.

Portanto:

$$\sigma^2\,(X \pm Y) = \sigma^2(X) + \sigma^2(Y)$$

Exemplo: Seja X uma variável aleatória contínua com a seguinte função densidade:

$f(x) = 0$	para	$x < 0$
$f(x) = 3x^2$	para	$0 < x < 1$
$f(x) = 0$	para	$x > 1$

Calcular: $E[X]$, Var $[X]$ e Desvio-padrão

$$E[X] = \int_{-\infty}^{\infty} x\,f(x)\,dx = \int_0^1 x \cdot 3x^2\,dx = 3\int_0^1 x^3\,dx =$$

$$= \frac{3x^4}{4}\Bigg|_0^1 = \frac{3}{4}$$

$$Var[X] = \int_{-\infty}^{\infty} (x_i - \mu_x)^2 f(x)\, dx = \int_o^1 \left(X - \frac{3}{4}\right)^2 3x^2\, dx = \frac{3}{80}$$

Desvio-padrão: $\sigma_{(x)} \sqrt{\frac{3}{80}} = 0{,}19$

2.14 COVARIÂNCIA E COEFICIENTE DE CORRELAÇÃO

O grau de dispersão conjunta e de associação linear entre duas variáveis aleatórias podem ser avaliados pela covariância e coeficiente de correlação.

2.14.1 Covariância

Define-se covariância entre x e y como:

$$COV_{XY} = E[(X_i - \mu_x)\ (Y_j - \mu_y)]$$

Desenvolvendo-se os parênteses, obtem-se a fórmula prática:

$$\boxed{COV_{XY} = E[XY] - \mu_x \mu_y}$$

2.14.2 Coeficiente de Correlação

Define-se coeficiente de correlação entre x e y como:

$$\boxed{\rho_{XY} = \frac{E[XY] - \mu_x \mu_y}{\sigma_{(x)} \cdot \sigma_y}} \qquad -1 \leq \rho\, xy \leq 1$$

onde:

$$E[XY] = \sum_i \sum_j x_i \cdot y_j\, p\,(x_i,\, y_j) \qquad \text{para } (X,Y) \text{ discreta}$$

ou

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, y\, f(x\,,\, y)\, dx\, dy \qquad \text{para } (X,Y) \text{ contínua}$$

Exemplo: Seja (X,Y) uma variável aleatória bidimensional discreta com a seguinte distribuição conjunta:

x \ y	–3	2	4
1	0,1	0,2	0,2
3	0,3	0,1	0,1

Calcular COV_{XY} e ρ_{XY}.

Primeiramente convém encontrar as distribuições marginais de X e Y. (Lembre-se soma das linhas e das colunas).

X	1	3
$P(x)$	0,5	0,5

Y	–3	2	4
$P(y)$	0,4	0,3	0,3

Então:

$$\mu_{(x)} = 1(0,5) + 3(0,5) = 2,0$$

$$\mu_{(y)} = -3(0,4) + 2(0,3) + 4(0,3) = 0,6$$

$$E[X^2] = (1)^2 \cdot 0,5 + (3)^2 \cdot \sqrt{x}0,5 = 5,0$$

$$E[Y^2] = (-3)^2\,0,4 + (2)^2\,0,3 + (4)^2\,0,3 = 9,6$$

$$\sigma^2_{(x)} = 5,0 - (2,0)^2 = 1 \rightarrow \therefore \sigma_{(x)} = 1$$

$$\sigma^2_{(y)} = 9,6 - (0,6)^2 = 9,24 \therefore \sigma_{(y)} = 3,04$$

$$E[XY] = \sum X_i\, Y_j\, p(X_i,\, Y_j)$$

$$= 1(-3)0,1 + 1(2)0,2 + 1(4)0,2 + 3(-3)0,3 + 3(2)0,1 +$$

$$+ 3(4)0,1 = 0$$

Logo,

$$COV_{XY} = E[XY] - \mu_x \mu_y$$

$$= 0 - (2)(0,6) = -1,2$$

$$\rho_{XY} = \frac{E[XY] - \mu_x\mu_y}{\sigma_x\,\sigma_y} = \frac{-1,2}{(1)(3,04)} \doteq -0,39$$

EXERCÍCIOS – SÉRIE III – CAPÍTULO 2

1. Considere a seguinte distribuição conjunta de X e Y:

X \ Y	–2	–1	4	5
1	0,1	0,2	0	0,3
2	0,2	0,1	0,1	0

a) achar as distribuições marginais de X e Y;
b) calcular $E[X]$, $E[Y]$ e $E[XY]$;
c) calcular covariância entre X e Y;
d) calcular σ_x e σ_y;
e) calcular ρ_{xy};
f) as variáveis são independentes? por quê?

2. Sejam M e N duas variáveis aleatórias com as seguintes distribuições:

M	1	3
$P(M)$	0,6	0,4

N	5	10	12
$P(N)$	0,3	0,5	0,2

a) achar a distribuição conjunta de (M,N);
b) calcule $E[M]$ e $E[N]$;
c) calcular $\sigma_{(M)}$ e $\sigma_{(N)}$;
d) qual é o valor de ρ_{NM}? por quê?

3. Dada a seguinte função densidade conjunta de (X,Y):

$$f(x, y) = \begin{cases} 3x^2y + 3y^2x & \text{para } \begin{cases} 0 \leq x \leq 1 \\ 0 \leq y \leq 1 \end{cases} \\ 0 & \text{para outros valores} \end{cases}$$

a) determinar as funções densidades marginais de X e Y;
b) calcular $E[X]$ e $E[Y]$;
c) calcular $\sigma^2_{(x)}$ e $\sigma^2_{(y)}$;
d) calcular $P(0{,}5 \leq x \leq 0{,}75)$;
e) calcular o coeficiente de correlação entre X e Y.

4. Suponha que (X, Y) tenha a seguinte função densidade de probabilidade:

$$f(x, y) = \begin{cases} x + y & \text{para } \begin{cases} 0 < x < 1 \\ 0 < y < 1 \end{cases} \\ 0 & \text{para outros valores} \end{cases}$$

a) calcule as distribuições marginais de X e de Y;
b) calcule $E[Y]$;
c) calcule $E[X]$;
d) covariância entre X e Y.

5. Uma variável aleatória X, tem uma densidade de probabilidade dada por:

$f(x) = \dfrac{K}{x}$ no intervalo $1 \leqslant x \leqslant 2$, e

$f(x) = 0$, fora desse intervalo.

a) determine K;
b) determine $\mu_{(x)}$;
c) determine a mediana de X;
d) determinar a moda de X;
e) calcule a variância de X.

6. X é uma variável aleatória contínua, tal que $f(x) = Kx^2 - Kx^3$ para $0 \leqslant x \leqslant 1$ e $f(x) = 0$ para outros valores.

a) ache K;
b) calcule a esperança de X;
c) calcule a mediana de X;
d) determine a variância de X.

7. X é uma variável aleatória discreta, tal que a função repartição é dada por:

$F(-2) = 0{,}3$ $\quad F(0) = 0{,}5$ $\quad F(1) = 0{,}6$

$F(2) = 0{,}8$ $\quad F(5) = 1{,}0$

a) calcule a média de x;
b) calcule a moda de X;
c) $P(-1 \leqslant x \leqslant 4)$?;
d) calcule a variância.

8. X é uma variável aleatória tal que a função repartição é dada por:

$F(x) = 0$ para $x < 0$

$F(x) = x^3$ para $0 \leq x < 1$

$F(x) = 1$ para $x \geq 1$

a) calcule a média;
b) determine a mediana;
c) calcule a variância.

9. Mostre que $COV_{XY} = E[XY] - \mu_x \mu_y$

10. Determinar a média e o desvio-padrão do peso líquido de um produto, sabendo-se que a média do peso bruto é 800 g, com desvio de 20 g e o peso da embalagem tem peso médio de 100 g, com desvio de 10 g.

11. A função de probabilidade conjunta de duas variáveis aleatórias discretas X e Y é dada por $P(x, y) = cxy$ para $x = 1, 2, 3$ e $y = 1, 2$, e zero em todos os outros casos. Determine:

a) o valor das constante c;
b) a tabela de distribuição conjunta de (X, Y);
c) $P(X = 3, Y = 2)$;
d) $F(2,2)$;
e) $P(X = 1)$;
f) $P(Y = 2)$;
g) a média de X;
h) a variância de Y;
i) $E[XY]$;
j) a covariância entre X e Y;
l) o coeficiente de correlação.

12. (X, Y) é uma variável bidimensional contínua com a função densidade conjunta:

$$f(x, y) = \begin{cases} c(x^2 + y^2) & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0, \text{ em caso contrário} \end{cases}$$

Determine:

a) a constante c;
b) $P(X < 1/2, Y > 1/2)$;

c) $P\left(\frac{1}{4} < x < \frac{3}{4}\right)$;

d) $P\left(Y < \frac{1}{2}\right)$;

e) se X e Y são independentes;

f) a covariância entre X e Y;

g) o coeficiência de correlação entre Y e X.

13. Um jogo consiste em se atirar um dado; se der faces dois ou cinco, a pessoa ganha \$ 50,00 por ponto obtido; se der faces um ou seis, a pessoa ganha \$ 100,00 por ponto obtido; se der faces três ou quatro, a pessoa paga \$ 150,00 por ponto obtido. Responda: O jogo é honesto? Calcule o desvio-padrão da distribuição.

14. Em uma classe, há 6 homens e 3 mulheres. Sorteados 3 alunos ao acaso e sem reposição, faça X: V.A. número de homens sorteados. Calcule a média, a moda e o desvio-padrão da distribuição.

15. Um processo de fabricação produz peças com peso médio de 30 g e desvio-padrão de 0,7 g. Essas peças são acondicionadas em pacotes de uma dezena cada. A embalagem pesa em média 40 g, com variância 2,25 g^2. Qual a média e o desvio-padrão do peso total do pacote?

16. O lucro unitário (L) de um produto é dado por $L = 1{,}2V - 0{,}8C - 3{,}5$. Sabendo-se que o preço unitário de venda (V) tem média \$ 60,00 e desvio-padrão \$ 5,00, e que o preço do custo unitário (C) tem uma distribuição de média \$ 50,00 e o desvio-padrão para \$ 2,00, qual a média e o desvio-padrão do lucro unitário?

3

MODELOS DE DISTRIBUIÇÕES DISCRETAS DE PROBABILIDADE

Neste capítulo são apresentados os principais modelos de distribuições de probabilidades para variáveis discretas.

3.1 DISTRIBUIÇÃO DE "BERNOULLI"

Suponhamos a realização de um experimento E, cujo resultado pode ser um sucesso (se acontecer o evento que nos interessa) ou um fracasso (o evento não se realiza).

Seja x a variável aleatória: sucesso ou fracasso.

$X \rightarrow x_1 = 1$ (sucesso) ou $x_2 = 0$ (fracasso)

$P(X) \rightarrow p(x_1) = p$ $\qquad p(x_2) = 1 - p = q.$

Diz-se que esta variável, assim definida, tem uma distribuição de "Bernoulli".

Suas principais características:

Média: $$\mu_{(X)} = \sum_{0}^{1} x_i \, P(x_i) = 0 \cdot q + 1 \cdot p = p$$

Variância: $$\sigma^2_{(X)} = E[(X_1 - \mu)^2] = E(X_i^2) - \mu^2_{(x)}$$

$$E[X_i^2] = \sum_{0}^{1} x_i^2 \, P(x_i) = 0^2 q + 1^2 p = p$$

$$\sigma^2_{(X)} = p - p^2 = p(1 - p) = pq$$

3.2 DISTRIBUIÇÃO BINOMIAL

Trata-se de uma distribuição de probabilidade adequada aos experimentos que apresentam apenas dois resultados (sucesso ou fracasso). Este modelo fundamenta-se nas seguintes hipóteses:

H1. n provas independentes e do mesmo tipo são realizadas;

H2. cada prova admite dois resultados – sucesso ou fracasso;

H3. a probabilidade de sucesso em cada prova é p e de fracasso $1 - p = q$.

Define-se a variável Y como o número de sucessos das n provas.

Logo, Y pode tomar os valores 0, 1, 2, 3, ..., n.

Fazendo sucesso corresponder a 1 e fracasso a 0, ou seja, provas de Bernoulli, tem-se:

Para $Y = 0$, uma seqüência de n zeros: 0000 ... 0. Logo:

$$P(Y = 0) = q \cdot q \cdot q \cdot q \dots q = q^n$$

Para $Y = 1$, uma seqüência do tipo: 1000 ... 0; 0100 ... 0; 001000 ... 0; serão n seqüências, cada uma com um único sucesso e $n - 1$ fracassos:

$$P(Y = 1) = n \cdot p \cdot q^{n-1}$$

Para $Y = y$, tem-se y sucessos e $(n - y)$ fracassos, correspondendo às seqüências com y algarismos 1 e $n - y$ zeros. Cada seqüência terá probabilidade $p^y q^{n-y}$ e como há $\binom{n}{y}$ seqüências distintas, tem-se:

$$P(Y = y) = \binom{n}{y} p^y q^{n-y}$$

que é a expressão geral da distribuição Binomial.

Para $Y = n$, tem-se uma seqüência de n uns: 1111 ... 1, logo:

$P(Y = n) = p^n$

O nome Binomial é porque $\binom{n}{y} p^y q^{n-y}$, nada mais é que o termo de grau y em p no desenvolvimento do Binômio de Newton $(q + p)^n$.

Média De acordo com as hipóteses, vê-se que Y é a soma de n variáveis do tipo "Bernoulli", daí,

$\mu_{(Y)} = n\,\mu_{(X)} = n \cdot p$ ou seja $\boxed{\mu_{(y)} = np}$

Variância Lembrando o que foi feito acima, temos:

$\sigma^2_{(Y)} = n\,\sigma^2_{(X)} = n \cdot p \cdot q$ ou seja $\boxed{\sigma^2_{(Y)} = npq}$

Exemplo: Uma moeda não viciada é lançada 8 vezes. Encontre a probabilidade de:

a) dar 5 caras;

b) pelo menos uma cara;

c) no máximo 2 caras.

Calcular a média e variância da distribuição.

Solução: Sabe-se que: $n = 8$, $p = \frac{1}{2}$ e $q = \frac{1}{2}$

$Y \rightarrow$ número de caras (sucessos)

a) $P(Y = 5) = \binom{8}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{8-5} = \frac{7}{32} = 0{,}22 = 22\%$

b) $P(Y \geq 1) = 1 - P(Y = 0) = 1 - \left(\frac{1}{2}\right)^8 = \frac{255}{256} = 0{,}996 = 99{,}6\%$

c) $P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) =$

$$= \left(\frac{1}{2}\right)^8 + 8\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^7 + \binom{8}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^6 =$$

$$= \frac{1}{256} + \frac{8}{256} + \frac{28}{256} = \frac{37}{256} = 0{,}14 = 14\%$$

A média será: $\mu_{(y)} = n \cdot p = 8 \cdot \frac{1}{2} = 4$

A variância será: $\sigma^2_{(y)} = n \cdot p \cdot q = 8 \cdot \frac{1}{2} \cdot \frac{1}{2} = 2$

3.3 DISTRIBUIÇÃO MULTINOMIAL

Uma generalização da Binomial resulta a distribuição multinomial. Assim, considere a possibilidade de K alternativas, isto é, repartirmos o Espaço Amostral em K eventos $A_1, A_2, ..., A_K$, mutuamente exclusivos com probabilidades $p_1, p_2, ..., p_K$, tais que $p_1 + p_2 + ... + p_K = 1$. Então n provas, a probabilidade de que A_1 ocorra X_1 vezes, A_2 ocorra X_2 vezes ... A_K ocorra X_K vezes é igual a:

$$P(X_1, X_2, ..., X_K) = \frac{n!}{X_1!\,X_2!\,...\,X_K!}\, p_1^{X_1} \cdot p_2^{X_2} ... p_K^{X_K}$$

Exemplo: Um dado é lançado 10 vezes. Qual é a probabilidade de terem aparecido duas vezes o nº 2, duas vezes o nº 5, três vezes o nº 1 e uma vez os demais resultados?

Então: $n = 10 \qquad p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$

$A_i = \{\text{sair o nº } i\}, \; i = 1, 2, 3, 4, 5, 6$

$X_1 = 3;\ X_2 = 2;\ X_3 = 1;\ X_4 = 1;\ X_5 = 2;\ X_6 = 1$

$$\sum_{i=1}^{6} X_i = 10$$

$P(X_1 = 3;\ X_2 = 2;\ X_3 = 1;\ X_4 = 1;\ X_5 = 2;\ X_6 = 1) =$

$$= \frac{10!}{3!\,2!\,1!\,1!\,2!\,1!}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1 \cong 0{,}0025 = 0{,}25\%$$

3.4 DISTRIBUIÇÃO DE POISSON

Em muitos casos, conhece-se o número de sucessos, porém se torna difícil e, às vezes, sem sentido, determinar o número de fracassos ou o número total de provas. Por exemplo: Automóveis que passam numa esquina. Pode-se num determinado intervalo de tempo anotar o número de carros que passaram, porém, o número de carros que deixaram de passar pela esquina não poderá ser determinado. Da mesma forma, o número de emendas num rolo de fita colante. Pode-se determinar quantas emendas existem, porém não é possível contar quantas emendas não ocorreram.

Analisando os exemplos acima, verifica-se que há uma variável t e quando $t \to \infty$ a probabilidade tende a aumentar.

Para encontrar a expressão que dá a probabilidade de x sucessos num intervalo t, algumas hipóteses precisam ser admitidas:

H1. $P(X = 1, \Delta t) = \lambda \Delta t$

H2. $P(X > 1, \Delta t) = 0$

H3. $P(X = 0, \Delta t) = 1 - \lambda \Delta t$

(a probabilidade de um sucesso num intervalo Δt é proporcional à amplitude do intervalo)

H4. As ocorrências de sucessos em intervalos são independentes.

Tem-se que: $\Delta t = \dfrac{t}{n}$, logo: $P(X = 1, \Delta t) = \lambda \dfrac{t}{n}$

Para se encontrar a expressão que dá $P(X, t)$, ou seja, a probabilidade de X sucessos no intervalo t, pode-se calcular o limite de uma distribuição Binomial com parâmetros n e $\dfrac{\lambda t}{n}$. Assim:

$$P(x, t) = \lim_{n \to \infty} \binom{n}{x} \left(\frac{\lambda t}{n}\right)^x \left(1 - \frac{\lambda t}{n}\right)^{n-x} =$$

$$= \lim_{n \to \infty} \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!} \frac{(\lambda t)^x}{n^x} \cdot \left(1 - \frac{\lambda t}{n}\right)^n \left(1 - \frac{\lambda t}{n}\right)^{-x}$$

$$= \lim_{n \to \infty} \frac{n(n-1)(n-2)\ldots(n-x+1)}{n \cdot n \cdot n \ldots\ldots\ldots n} \cdot \frac{(\lambda t)^x}{x!} \left(1 - \frac{\lambda t}{n}\right)^n \left(1 - \frac{\lambda t}{n}\right)^{-x}$$

$$= \lim_{n \to \infty} \underbrace{\left[1\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{x+1}{n}\right)\left(1 - \frac{\lambda t}{n}\right)^{-x}\right]}_{1} \underbrace{\left(1 - \frac{\lambda t}{n}\right)^n}_{e^{-\lambda t}} \cdot \frac{\lambda t^x}{x!}$$

Portanto:

$$\boxed{P(x, t) = \frac{(\lambda t)^x}{x!} \cdot e^{-\lambda t}}$$

que é a fórmula para o cálculo de variáveis com distribuição de Poisson. Onde λ é o coeficiente de proporcionalidade e $e = 2, 7 \ldots$

Média Pode-se provar que $\boxed{\mu_{(x)} = \lambda t}$

Variância Pode-se provar que $\boxed{\sigma^2_{(x)} = \lambda t}$

Exemplo: Em média há 2 chamadas por hora num certo telefone. Calcular a probabilidade de se receber no máximo 3 chamadas em 2 horas e a probabilidade de nenhum chamada em 90 minutos.

Solução: $\mu = \lambda t : 2 = \lambda t \therefore \lambda = 2$

$$P(X \leq 3, 2h) = P(X = 0, 2) + P(X = 1, 2) + P(X = 2, 2) + P(X = 3, 2) =$$

$$= \frac{(4)^0 e^{-4}}{0!} + \frac{(4)^1 e^{-4}}{1!} + \frac{(4)^2 e^{-4}}{2!} + \frac{(4)^3 e^{-4}}{3!} =$$

$$= 0{,}0183 + 0{,}0732 + 0{,}1464 + 0{,}1952 = 0{,}4331 = 43{,}31\%$$

$\mu = \lambda t$

$$2 = \lambda 60 \therefore \lambda = \frac{2}{60} = \frac{1}{30}$$

$$P(X = 0{,}90) = \frac{\left(\frac{90}{30}\right)^0 e^{-\left(\frac{90}{30}\right)}}{0!} = e^{-3} = 0{,}0498 = 4{,}98\%$$

EXERCÍCIOS – SÉRIE I – CAPÍTULO 3

Distribuição Binomial

1. Uma moeda é jogada 10 vezes. Calcular as seguintes probabilidades:
   a) de ocorrer 6 caras;
   b) de dar pelo menos 2 caras;
   c) de não dar nenhuma coroa;
   d) de dar pelo menos uma coroa;
   e) de não dar 5 caras e 5 coroas.

2. Admitindo-se que os nascimentos de meninos e meninas sejam iguais, calcular a probabilidade de um casal com 6 filhos ter 4 filhos homens e 2 mulheres.

3. Em 320 famílias com 4 crianças cada uma, quantas se esperaria que tivessem:
   a) nenhuma menina;
   b) 3 meninos;
   c) 4 meninos.

4. Qual a probabilidade de obter ao menos uma vez o ponto 3 em *n* jogadas de um dado?

5. Um time X tem $\frac{2}{3}$ de probabilidade de vitória sempre que joga. Se X jogar 5 partidas, calcule a probabilidade de:

a) X vencer exatamente 3 partidas;
b) X vencer ao menos uma partida;
c) X vencer mais da metade das partidas.

6. A probabilidade de um atirador acertar o alvo é $\frac{1}{3}$. Se ele atirar 6 vezes, qual a probabilidade de:

a) acertar exatamente 2 tiros?
b) não acertar nenhum tiro?

7. Num teste do tipo certo-errado, com 100 perguntas, qual a probabilidade de um aluno, respondendo as questões ao acaso, acertar 70% das perguntas?

8. Uma variável aleatória com distribuição binomial tem a função repartição dada por:

$F_{(0)} = \frac{1}{243}$ $F_{(3)} = \frac{131}{243}$

$F_{(1)} = \frac{11}{243}$ $F_{(4)} = \frac{211}{243}$

$F_{(2)} = \frac{51}{243}$ $F_{(5)} = 1$

Determinar:

a) n
b) p e q
c) média de Y
d) variância de Y
e) $P(Y \geq 1)$
f) $P(2 \leq Y \leq 4)$

9. Se 5% das lâmpadas de certa marca são defeituosas, achar a probabilidade de que, numa amostra de 100 lâmpadas, escolhidas ao acaso, tenhamos:

a) nenhuma defeituosa;
b) 3 defeituosas;
c) mais do que 1 boa.

10. Aplique a definição de Média e Variância de uma Variável Aleatória Discreta para provar que a média de uma binomial é $n \cdot p$ e a variância $n \cdot p \cdot q$.

Distribuição Multinomial

11. Jogue um dado 8 vezes. Calcule a probabilidade de aparecer 2 números 2; 2 números 5 e os demais números, uma vez.

12. As lâmpadas coloridas produzidas por uma fábrica são 60% verdes, 30% azuis e 10% amarelas. Em 5 lâmpadas, encontre a probabilidade de que 2 sejam verdes, 1 azul e 2 amarelas.

13. O sangue humano foi classificado em 4 tipos: A, O, B e AB. Numa certa população, as probabilidades destes tipos são respectivamente: 0,40; 0,45; 0,10 e 0,05. Qual a probabilidade de que em 5 indivíduos escolhidos ao acaso haja:

    a) dois do tipo A e um de cada um dos outros?

    b) três do tipo A e dois do tipo O?

Distribuição de Poisson

14. Uma fábrica de pneus verificou que ao testar seus pneus nas pistas, havia em média um estouro de pneu a cada 5.000 km.

    a) Qual a probabilidade que num teste de 3.000 km haja no máximo um pneu estourado?

    b) Qual a probabilidade de que um carro ande 8.000 km sem estourar nenhum pneu?

15. Certo posto de bombeiros recebe em média 3 chamadas por dia. Calcular a probabilidade de:

    a) receber 4 chamadas num dia;

    b) receber 3 ou mais chamadas num dia.

16. A média de chamadas telefônicas numa hora é 3. Qual a probabilidade de:

    a) receber exatamente 3 chamadas numa hora?

    b) receber 4 ou mais chamadas em 90 minutos?

17. Na pintura de paredes aparecem defeitos em média na proporção de 1 defeito por metro quadrado. Qual a probabilidade de aparecerem 3 defeitos numa parede de 2×2m?

18. Suponha que haja em média 2 suicídios por ano numa população de 50.000. Em uma cidade de 100.000 habitantes, encontre a probabilidade de que em um dado ano tenha havido:

a) 0;

b) 1;

c) 2;

d) 2 ou mais suicídios.

19. Suponha 400 erros de impressão distribuídos aleatoriamente em um livro de 500 páginas. Encontre a probabilidade de que uma dada página contenha:

a) nenhum erro;

b) exatamente 2 erros.

20. Uma loja atende em média 2 clientes por hora. Calcular a probabilidade de em uma hora:

a) atender exatamente 2 clientes;

b) atender 3 clientes.

21. Aplicando as definições de média e variância, prove que a média e a variância de uma Poisson são iguais a λt.

4

Modelos de Distribuições Contínuas de Probabilidade

Os principais modelos de distribuições de variáveis aleatórias contínuas são apresentados neste capítulo.

4.1 DISTRIBUIÇÃO UNIFORME OU RETANGULAR

X é uma variável uniformemente distribuída no intervalo $[a, b]$ se sua função densidade for dada por:

$$f(x) = 0 \quad \text{para} \quad x \text{ fora de } [a, b]$$

$$f(x) = \frac{1}{b - a} \quad \text{para} \quad a \leq x \leq b$$

Sua função repartição é:

$$F(x) = 0 \quad \text{para} \quad x < a$$

$$F(x) = \frac{x - a}{b - a} \quad \text{para} \quad a \leq x < b$$

$$F(x) = 1 \quad \text{para} \quad x \geq b$$

Pode-se demonstrar que a média e a variância são dadas por:

$$\mu(x) = \frac{a + b}{2}$$

$$\sigma^2_{(x)} = \frac{(b - a)^2}{12}$$

Exemplo: Um ponto é escolhido ao acaso no segmento de reta [0, 2]. Qual será a probabilidade de que o ponto escolhido esteja entre 1 e $\frac{3}{2}$?

Fazendo X representar a variável escolher um ponto de [0, 2] tem-se que a função densidade de X é dada por $f(x) = \frac{1}{b-a} = \frac{1}{2-0} = \frac{1}{2}$, para $0 \leq x \leq 2$

$$P\left(1 \leq x \leq \frac{3}{2}\right) = \int_1^{1,5} \frac{1}{2}\, dx = \frac{1}{4}.$$

4.2 DISTRIBUIÇÃO NORMAL

É a mais importante distribuição de probabilidade, sendo aplicada em inúmeros fenômenos e utilizada para o desenvolvimento teórico da estatística. É também conhecida como distribuição de Gauss, Laplace ou Laplace-Gauss.

Seja X uma variável aleatória contínua. X terá distribuição normal se:

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; -\infty < x < \infty$$

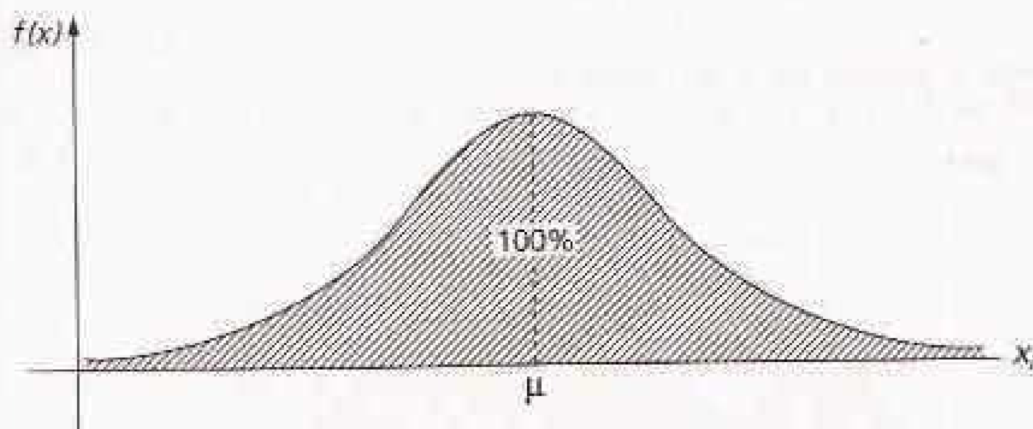
onde: μ = média de distribuição

σ = desvio-padrão da distribuição

π = 3,1416 ...

e = 2,7 ...

Sendo seu gráfico:



Para o cálculo das probabilidades, surgem dois grandes problemas: primeiro, para a integração de $f(x)$, pois para o cálculo é necessário o desen-

volvimento em séries; segundo, seria a elaboração de uma tabela de probabilidades, pois $f(x)$ depende de dois parâmetros, fato este que acarretaria um grande trabalho para tabelar essas probabilidades considerando-se as várias combinações de μ e σ^2.

Esses problemas podem ser solucionados por meio de uma mudança de variável, obtendo-se, assim, a distribuição normal padronizada ou reduzida.

4.2.1 Distribuição Normal Padrão

Seja Z uma variável aleatória tal que:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

em que X é uma variável normal de média μ e variância σ^2.

Então a média de z será:

$$E[z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\,|E[X-\mu]| = \frac{1}{\sigma}\,|E[X] - E[\mu]| = \frac{1}{\sigma}\,|\mu - \mu| = 0$$

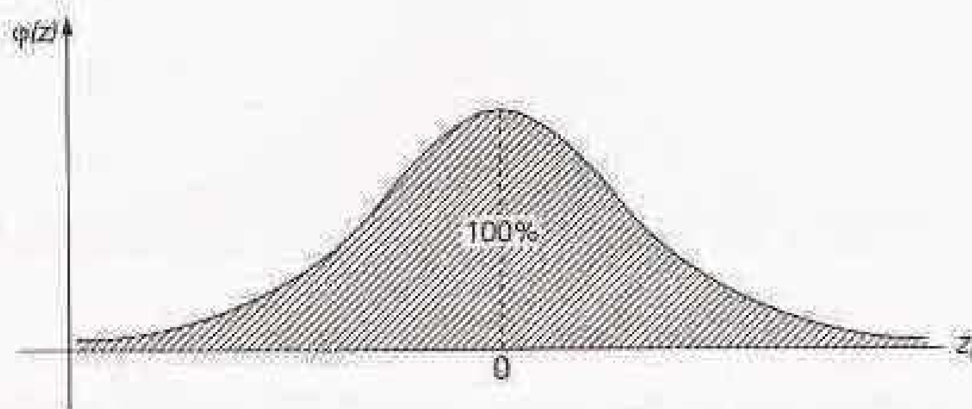
E sua variância:

$$\text{Var}[z] = \text{Var}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma^2}\,|\text{Var}[X-\mu]| = \frac{1}{\sigma^2}\,|\text{Var}[X]| = \frac{\sigma^2}{\sigma^2} = 1$$

Logo, a função densidade será:

$$\varphi(z) = \frac{1}{\sqrt{2\Pi}}\, e^{-\frac{1}{2}z^2} \quad , -\infty < z < \infty$$

Sendo o gráfico de $\varphi(z)$ igual a:



Como a média de Z é 0 e a variância 1, as probabilidades (áreas) sob $\varphi(z)$ são calculadas e tabeladas. Nos exemplos seguintes será explicado o uso da tabela da distribuição normal padrão.

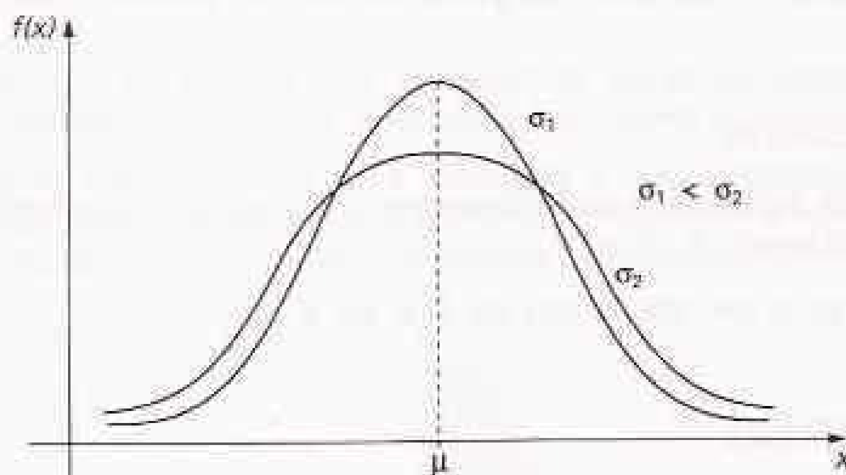
Para se registrarem distribuições normais usa-se a seguinte notação:

$X \overset{d}{=} N(\mu\,;\sigma^2)$ lê-se "a variável aleatória X tem distribuição normal com média μ e variância σ^2."

$Z \overset{d}{=} N(0\,;1)$ lê-se "a variável aleatória Z tem distribuição normal com média 0 e variância 1." Ou, simplesmente, distribuição normal padrão.

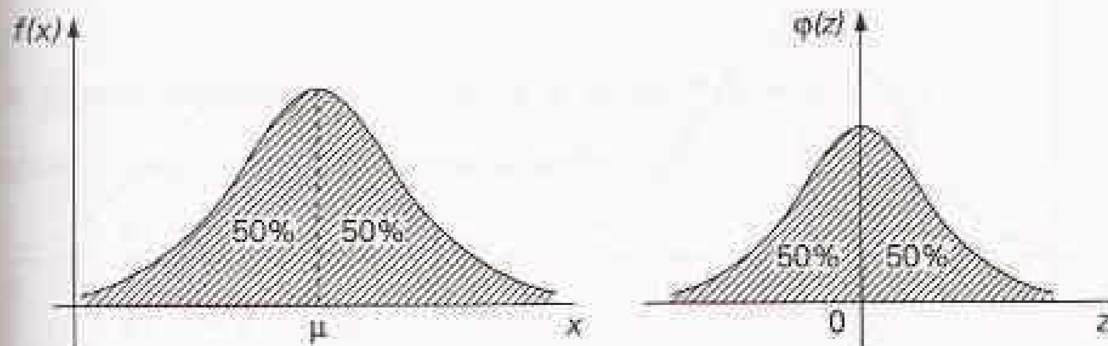
4.2.2 Propriedades da Distribuição Normal

Como foi visto, o gráfico da função densidade de uma variável normal tem a forma de um sino e é simétrico em relação à média μ. Fixando-se a média, verifica-se que o "achatamento" está diretamente ligado ao valor de σ; assim:



1ª PROPRIEDADE

"$f(x)$ é simétrica em relação à origem $x = \mu$, ou $\varphi(z)$ é simétrica em relação à origem $Z = 0$.

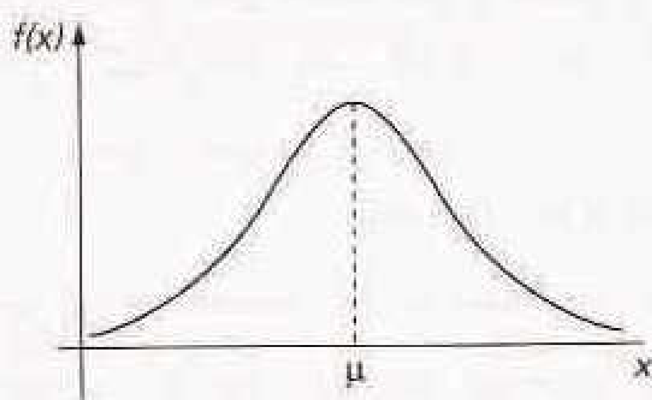


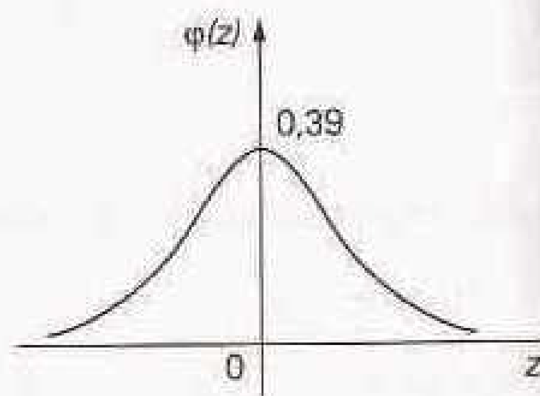
2ª PROPRIEDADE

"$f(x)$ possui um máximo para

$x = \mu$, ou $\varphi(z)$ possui um máximo para

$Z = 0$, e neste caso sua ordenada vale 0,39"





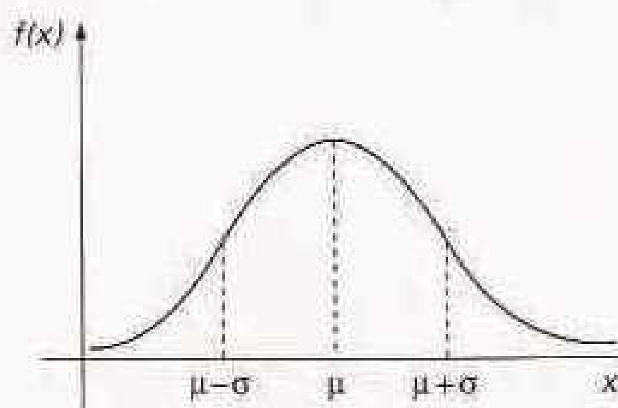
3ª PROPRIEDADE

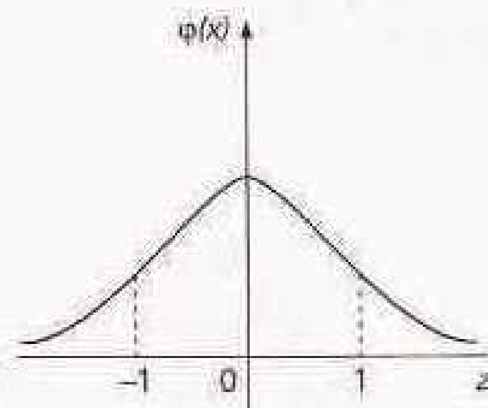
"$f(x)$ tende a zero quando x tende para $+\infty$ ou $-\infty$, o mesmo acontecendo com $\varphi(z)$ quando $Z \rightarrow \pm\infty$.

Isto é, x ou z são assíntotas de $f(x)$ ou $\varphi(z)$.

4ª PROPRIEDADE

"$f(x)$ tem dois pontos de inflexão cujas abscissas valem $\mu + \sigma$ e $\mu - \sigma$, ou $\varphi(z)$ tem dois pontos de inflexão cujas abscissas valem $+1$ e -1."





4.2.3 Combinação de Distribuições Normais

"A combinação linear de variáveis com distribuições normais independentes é também variável com distribuição normal."

Por exemplo: se X e Y são variáveis aleatórias com distribuições normais, então:

$$W = aX + bY + c$$

terá distribuição normal com:

$$\mu(w) = a \cdot \mu(x) + b \cdot \mu(y) + c, \text{ e}$$

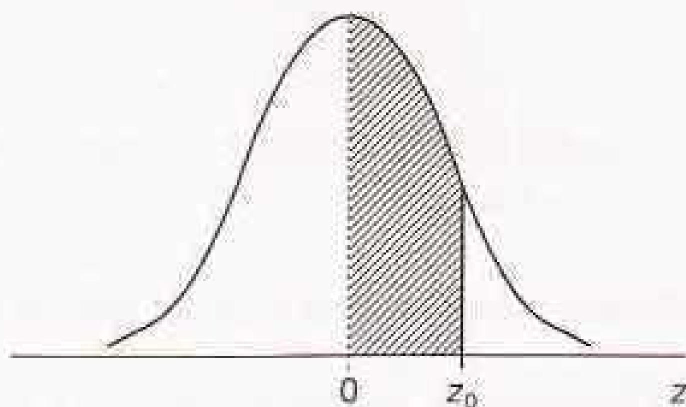
$$\sigma^2(w) = a^2 \cdot \sigma^2(x) + b^2 \cdot \sigma^2(y)$$

Isto é, em particular, a soma ou diferença de duas variáveis aleatórias normais também é uma variável aleatória normal.

4.2.4 Uso da Tabela de Distribuição Normal Padrão

Há vários tipos de tabelas que oferecem as áreas (probabilidades) sob a curva normal padrão. O tipo mais freqüente é a Tabela de Faixa Central.

A Tabela de Faixa Central dá a área sob a curva normal padrão entre $z = 0$ e qualquer valor positivo de z. A simetria em torno de $z = 0$ permite obter a área entre quaisquer valores de z (positivos ou negativos).



A Tabela oferece a área entre 0 e z_0 ou $P(0 \leq z \leq z_0)$.

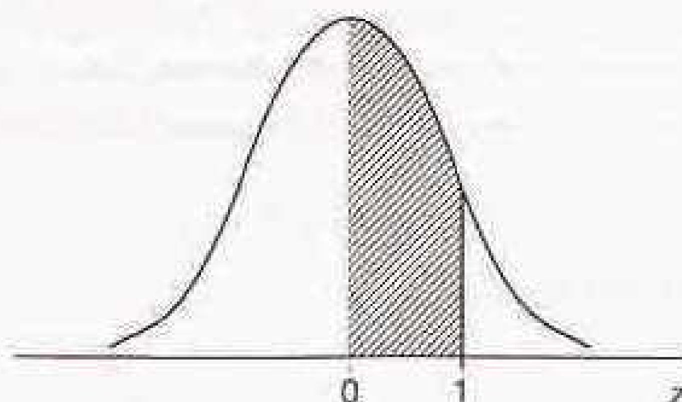
Exemplo: Desejam-se as probabilidades:

a) $P(0 \leq z \leq 1)$

b) $P(-2{,}55 < z < 1{,}2)$

c) $P(z \geq 1{,}93)$

No caso a) *tem-se:*

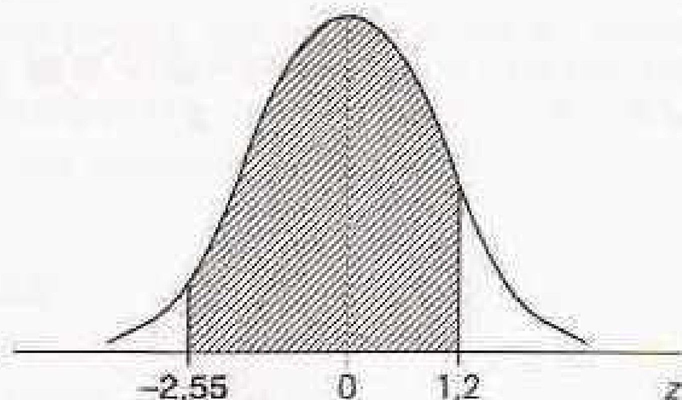


Para se obter probabilidade, basta entrar com a abscissa 1,0 (na primeira coluna) e 0,00 (na primeira linha) da Tabela. Assim:

$$P(0 \leq z \leq 1) = 0{,}3413$$

Confirme o resultado, consultando a Tabela 1 no final do livro.

No caso b)

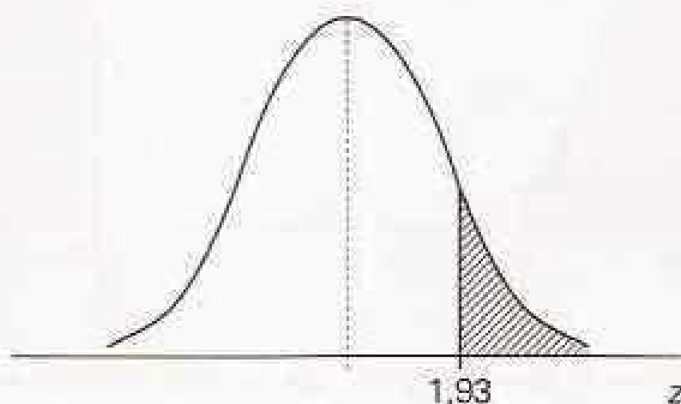


Entra-se, na Tabela, com o valor 1,2 na 1ª coluna e 00 na 1ª linha, obtendo 0,3849.

Lembrando a propriedade da simetria em relação a $z = 0$, entra-se com 2,5 na 1ª coluna e 05 na primeira linha, obtendo 0,4946. Portanto,

$$P(-2{,}55 < z < 1{,}2) = 0{,}3849 + 0{,}4946 = 0{,}8795$$

No caso c)



Entra-se na Tabela com 1,9 na 1ª coluna e 0,03 na 1ª linha obtendo 0,4732. Porém, essa é a área compreendida entre 0 e 1,93. Lembrando que a área embaixo da curva vale 1 e que a função é simétrica em relação à origem $z = 0$, tem-se

$$P(z > 1{,}93) = 0{,}5000 - 0{,}4732 = 0{,}0268.$$

Exemplos:

1. As alturas dos alunos de determinada escola são normalmente distribuídas com média 1,60 m e desvio-padrão 0,30 m. Encontre a probabilidade de um aluno medir:
   a) entre 1,50 e 1,80 m;
   b) mais de 1,75 m;
   c) menos de 1,48 m.
   Qual deve ser a medida mínima para escolhermos 10% dos mais altos?

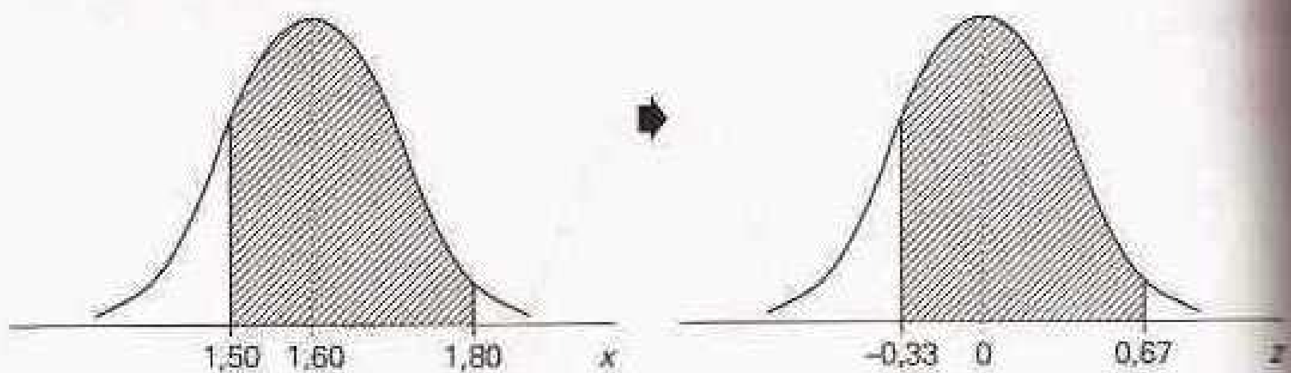
Solução: Sabe-se que $\mu = 1{,}60$ e $\sigma = 0{,}30$.

Faça X a variável normal altura dos alunos. Então:

a) $P(1{,}50 \leq X \leq 1{,}80) = P(z_1 \leq z \leq z_2) =$
$= P(-0{,}33 \leq z \leq 0{,}67) =$
$= 0{,}1293 + 0{,}2486 = 0{,}3779 = 37{,}79\%.$

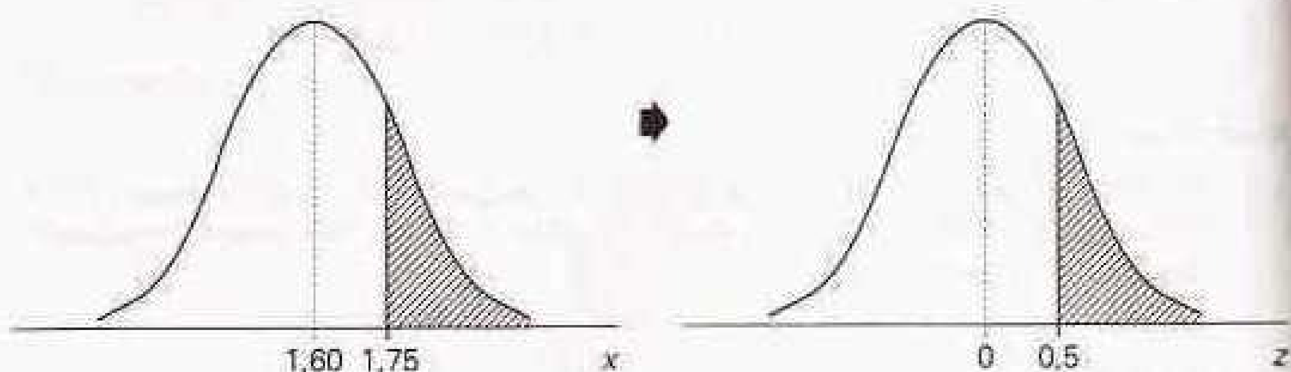
em que $z_1 = \dfrac{X - \mu}{\sigma} = \dfrac{1{,}50 - 1{,}60}{0{,}30} = -0{,}33$, e

$$z_2 = \frac{X - \mu}{\sigma} = \frac{1{,}80 - 1{,}60}{0{,}30} = 0{,}67$$



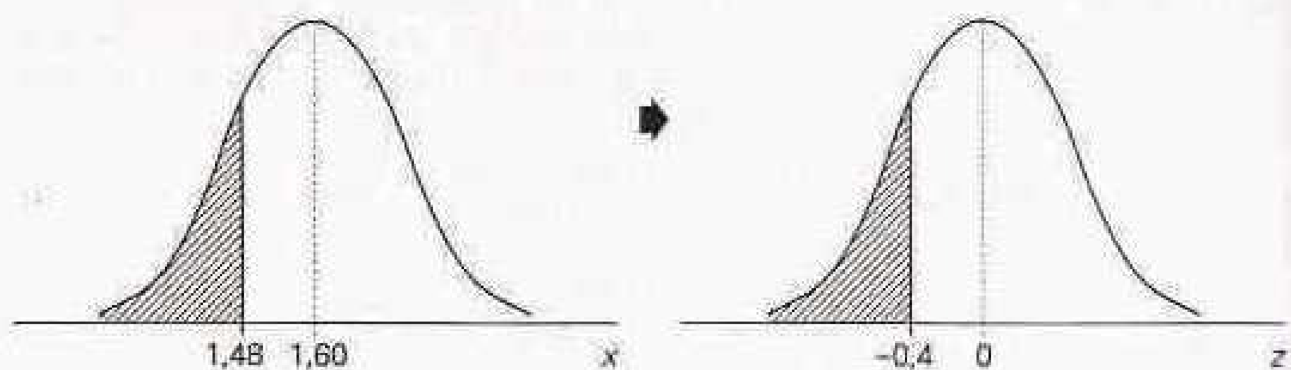
b) $P(X > 1,75) = P(z > z_1) = P(z > 0,5) =$
$= 0,5000 - 0,1915 = 0,3085$

em que $z_1 = \dfrac{1,75 - 1,60}{0,30} = 0,5$

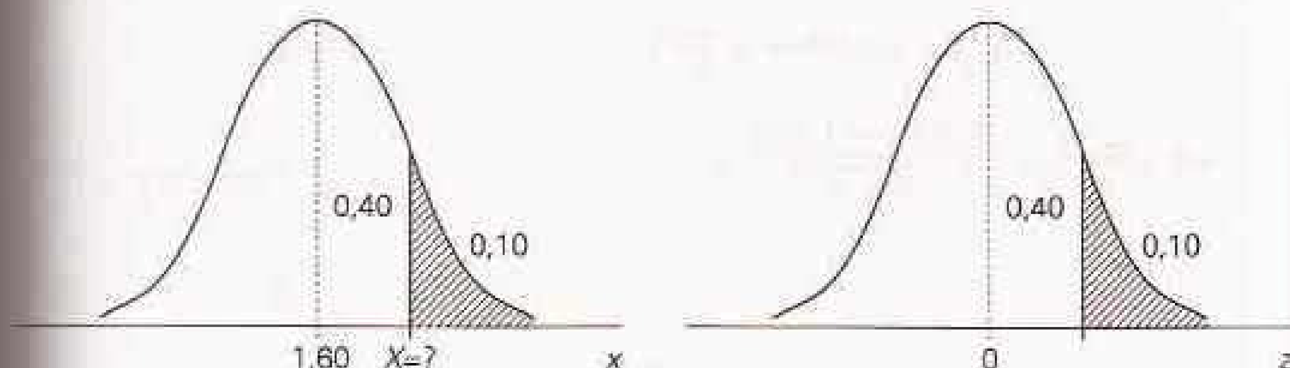


c) $P(X < 1,48) = P(z < z_1) = P(z < -0,4) =$
$= 0,5000 - 0,1554 = 0,3446$

em que $z_1 = \dfrac{1,48 - 1,50}{0,30} \cong -0,4$



d) É o problema inverso dos itens anteriores, pois neste caso tem-se a probabilidade e deseja-se a medida:



Para se encontrar o valor de z que deixa 0,10 à direita, deve-se entrar na tabela com 0,40, assim: descobrimos que $z = 1{,}28$. Logo,

$$z = \frac{X - \mu}{\sigma} \Rightarrow 1{,}28 = \frac{X - 1{,}60}{0{,}30}$$

portanto $X = 1{,}98$ m deve ser a medida para se encontrar 10% dos mais altos.

2. Um produto pesa, em média, 10 g, com desvio-padrão de 2 g. É embalado em caixas com 50 unidades. Sabe-se que as caixas vazias pesam 500 g, com desvio-padrão de 25 g. Admitindo-se uma distribuição normal dos pesos e independência entre as variáveis dos pesos do produto e da caixa, calcular a probabilidade de uma caixa cheia pesar mais de 1.050 g.

Solução: Peso do Produto: $\mu_p = 10$ e $\sigma_p = 2$

Peso da Caixa: $\mu_c = 500$ e $\sigma_c = 25$

Então, a média da caixa cheia será:

$$\mu_{total} = 50 \cdot 10 + 500 = 1.000$$

Quanto ao desvio-padrão da caixa cheia, deve-se lembrar as propriedades da variância. Assim:

$$\sigma^2_{total} = 50\,(\text{ Var. do produto}) + \text{Var. da Caixa:}$$

$$= 50(2)^2 + (25)^2 = 825$$

Portanto, o desvio-padrão do total será:

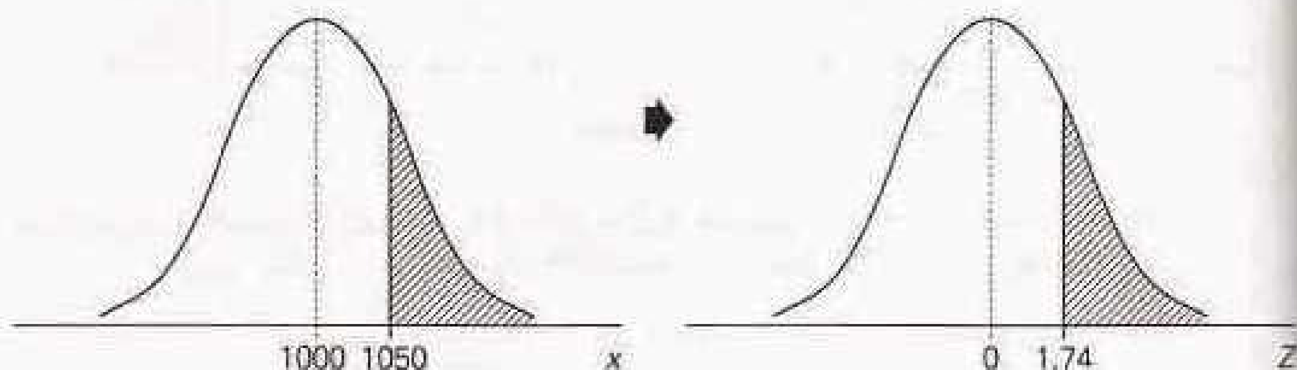
$$\sigma_{total} = \sqrt{825} \cong 28{,}72$$

Então,

$$P(X > 1.050) = P(z > z_1) = P(z > 1{,}74) = 0{,}5000 - 0{,}4591 =$$

$$= 0{,}0409 = 4{,}09\%$$

em que $z_1 = \dfrac{1.050 - 1.000}{28{,}72} = 1{,}74$



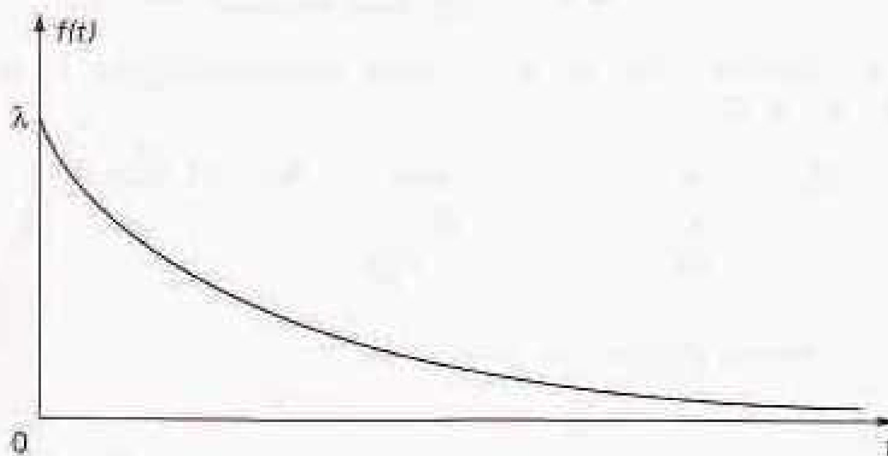
4.3 DISTRIBUIÇÃO EXPONENCIAL

A distribuição de probabilidade do intervalo t entre dois sucessos consecutivos de uma lei de Poisson é a distribuição exponencial.

Sua função densidade é dada por:

$$\begin{cases} f(t) = 0 & \text{para } t < 0 \\ f(t) = \lambda e^{-\lambda t} & \text{para } t \geq 0, \lambda > 0 \end{cases}$$

O gráfico de $f(t)$ é dado por:

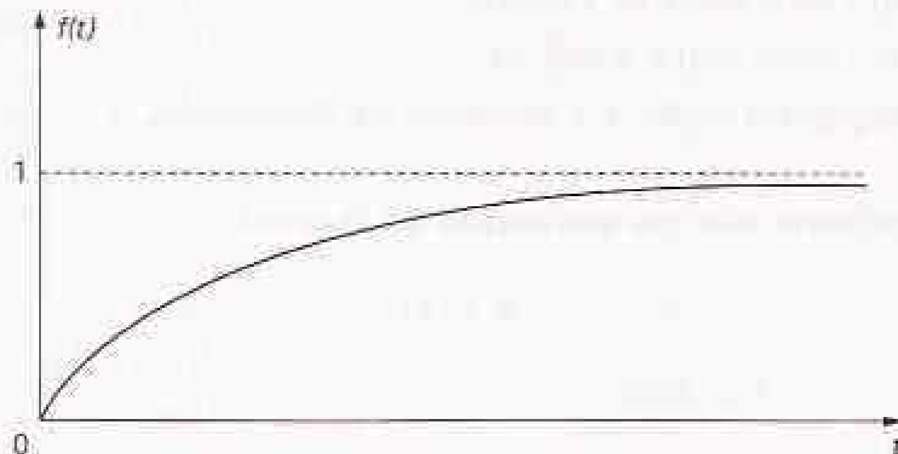


4.3.1 Função Repartição:

para $t < 0$; $F(t) = 0$

para $t \geqslant 0$; $F(t) = \int_0^t \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t}$

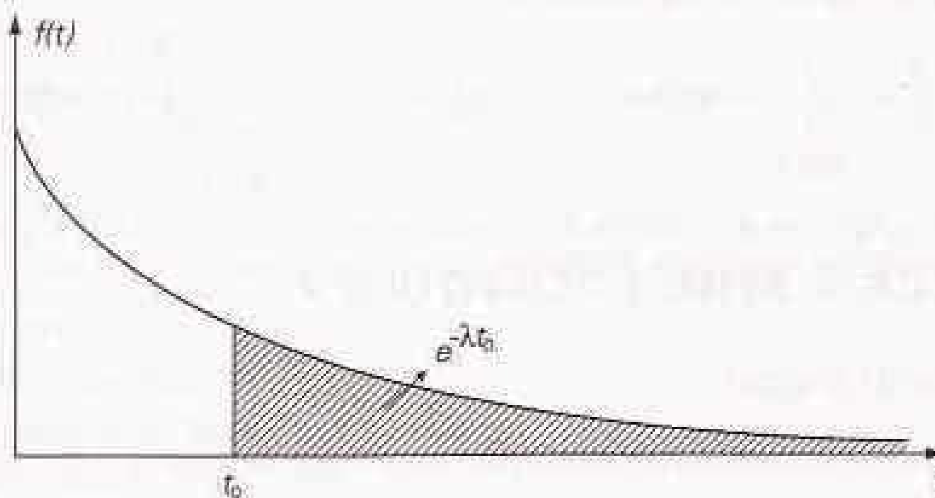
Gráfico de $F(t)$



Conhecida a função repartição de t, pode-se facilmente determinar:

$$P(T \geqslant t_0) = 1 - F(t_0) = 1 - [1 - e^{-\lambda t_0}] = e^{-\lambda t_0}$$

Veja o gráfico:



Quanto à média: $\mu_{(t)} = \dfrac{1}{\lambda}$

Quanto à variância: $\sigma^2_{(t)} = \dfrac{1}{\lambda^2}$

Exemplo: Os defeitos de um tecido seguem a distribuição de Poisson com média de um defeito a cada 400 m. Qual a probabilidade de que o intervalo entre dois defeitos consecutivos seja:

a) no mínimo de 1.000 m;

b) entre 800 e 1.000 m.

Calcule a média e a variância da distribuição.

Solução: Sabe-se que na distribuição de Poisson,

$$\mu = \lambda t$$

logo $1 = \lambda 400$ $\lambda = \dfrac{1}{400}$

a) $P(t \geq 1.000) = e^{-\lambda t_0} = e^{-\frac{1.000}{400}} = e^{-2,5} = 0,0821 = 8,21\%$

b) $P(800 \leq t \leq 1.000) = P(t \geq 800) - P(t \geq 1.000)$

$$= e^{-\frac{800}{400}}\ e^{-\frac{1.000}{400}} = e^{-2} - e^{-2,5} = 0,0532 = 5,32\%$$

Quanto à média e à variância de t:

$$\mu_{(t)} = \frac{1}{\lambda} = \frac{1}{\frac{1}{400}} = 400\text{ m} \qquad \sigma^2_{(t)} = \frac{1}{\lambda^2} = \frac{1}{\left(\frac{1}{400}\right)^2} = 160000\text{ m}^2$$

EXERCÍCIOS – SÉRIE I – CAPÍTULO 4

Distribuição Uniforme

1. Um ponto é escolhido ao acaso no segmento de reta [1, 4]. Calcular:
   a) probabilidade de que o ponto escolhido esteja entre 2 e 3;
   b) entre 0,5 e 2,5;
   c) seja exatamente o 2;

d) a média dessa distribuição;
e) a variância dessa distribuição.

2. Calcular a expressão para a média e variância de uma variável uniformemente distribuída entre a e b.

3. Suponha que X seja uniformemente distribuído entre $[-\alpha, \alpha]$, em que $\alpha > 0$. Quando possível, calcular α de modo que as seguintes relações sejam satisfeitas:

a) $P(X > 1) = \frac{1}{3}$

b) $P(X > 1) = \frac{1}{2}$

c) $P\left(X < \frac{1}{2}\right) = 0{,}7$

d) $P\left(X < \frac{1}{2}\right) = 0{,}3$

Distribuição Normal

4. Faça z uma variável com distribuição normal padronizada e encontre (use a tabela):

a) $P(0 \leq z \leq 1{,}44)$
b) $P(-0{,}85 < z < 0)$
c) $P(-1{,}48 < z < 2{,}05)$
d) $P(0{,}72 < z < 1{,}89)$
e) $P(z \geq 1{,}08)$
f) $P(z \geq -0{,}66)$
g) $P(|z| \leq 0{,}5)$

5. A duração de um certo componente eletrônico tem média 850 dias e desvio-padrão 45 dias. Calcular a probabilidade desse componente durar:

a) entre 700 e 1.000 dias;
b) mais que 800 dias;
c) menos que 750 dias;
d) exatamente 1.000 dias.

Qual deve ser o número de dias necessários para que tenhamos de repor no máximo 5% dos componentes?

6. Os pesos de 600 estudantes são normalmente distribuídos com média 65,3 kg e desvio-padrão 5,5 kg. Encontre o número de alunos que pesam:

   a) entre 60 e 70 kg;
   b) mais que 63,2 kg.

7. Suponha que as notas de uma prova sejam normalmente distribuídas com média 73 e desvio-padrão 15. 15% dos alunos mais adiantados recebem a nota A e 12% dos mais atrasados recebem nota F. Encontre o mínimo para receber A e o mínimo para passar, não receber F.

8. Uma fábrica de pneumáticos fez um teste para medir o desgaste de seus pneus e verificou que ele obedecia a uma distribuição normal, de média 48.000 km e desvio-padrão 2.000 km. Calcular a probabilidade de um pneu escolhido ao acaso:

   a) dure mais que 46.000 km;
   b) dure entre 45.000 e 50.000 km.

9. X é uma variável aleatória contínua, tal que $X = N(12,25)$. Qual a probabilidade de uma observação ao acaso:

   a) ser menor do que –3;
   b) cair entre –1 e 15.

10. O salário semanal dos operários industriais são distribuídos normalmente em torno de uma média de $ 180,00 com desvio-padrão de $ 25,00. Pede-se:

    a) encontre a probabilidade de um operário ter salário semanal situado entre $ 150,00 e $ 178,00;
    b) dentro de que desvios de ambos os lados da média, cairão 96% dos salários?

11. Certo produto tem peso médio de 10 g e desvio-padrão 0,5 g. É embalado em caixas de 120 unidades que pesam em média 150 g e desvio-padrão 8 g. Qual a probabilidade de que uma caixa cheia pese mais de 1.370 g?

12. Determinada máquina enche latas baseada no peso bruto com média 1 kg e desvio-padrão 25 g.

    As latas têm peso de 90 g com desvio-padrão 8 g. Pede-se:

    a) a probabilidade de uma lata conter menos de 870 g de peso líquido;
    b) a probabilidade de uma lata conter mais de 900 g de peso líquido.

13. Um avião de turismo de 4 lugares pode levar uma carga útil de 350 kg. Supondo que os passageiros têm peso de 70 kg com distribuição normal de peso e desvio-padrão 20 kg, e que a bagagem de cada passageiro

pese em média 12 kg, com desvio-padrão 5 kg e distribuição normal do peso. Calcular a probabilidade de:

a) haver sobrecarga se o piloto não pesar os 4 passageiros e respectiva bagagem;

b) que o piloto tenha de tirar pelo menos 50 kg de gasolina para evitar sobrecarga.

14. Em uma distribuição normal, 28% dos elementos são superiores a 34 e 12% inferiores a 19. Encontrar a média e a variância da distribuição.

15. Seja Y uma função tal que $Y = X_1 + X_2 + X_3$ e as variáveis X_i são independentes com as seguintes distribuições: $X_1 = N(10;9)$; $X_2 = N(-2;4)$; $X_3 = N(5;25)$. Qual é a distribuição de Y?

16. Suponha que o diâmetro médio dos parafusos produzidos por uma fábrica é de 0,25 polegadas, e o desvio-padrão 0,02 polegadas. Um parafuso é considerado defeituoso se seu diâmetro é maior que 0,28 polegadas ou menor que 0,20 polegadas.

a) Encontre a porcentagem de parafusos defeituosos;

b) Qual deve ser a medida mínima para que tenhamos no máximo 12% de parafusos defeituosos?

17. Suponha que a duração de vida de dois equipamentos E_1 e E_2 tenham respectivamente distribuições: $N(45;9)$ e $N(40;36)$. Se o equipamento tiver que ser usado por um período de 45 horas, qual deles deve ser preferido?

18. Certa máquina de empacotar determinado produto oferece variações de peso com desvio-padrão de 20 g. Em quanto deve ser regulado o peso médio do pacote para que apenas 10% tenham menos que 400 g?

Calcule a probabilidade de um pacote sair com mais de 450 g.

19. Sendo $X \sim N(\mu; \sigma^2)$, determine:

a) $P(\mu - \sigma \leq X \leq \mu + \sigma)$

b) $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$

c) $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$

d) $P(\mu - 1{,}5\sigma \leq X \leq \mu + 1{,}5\sigma)$

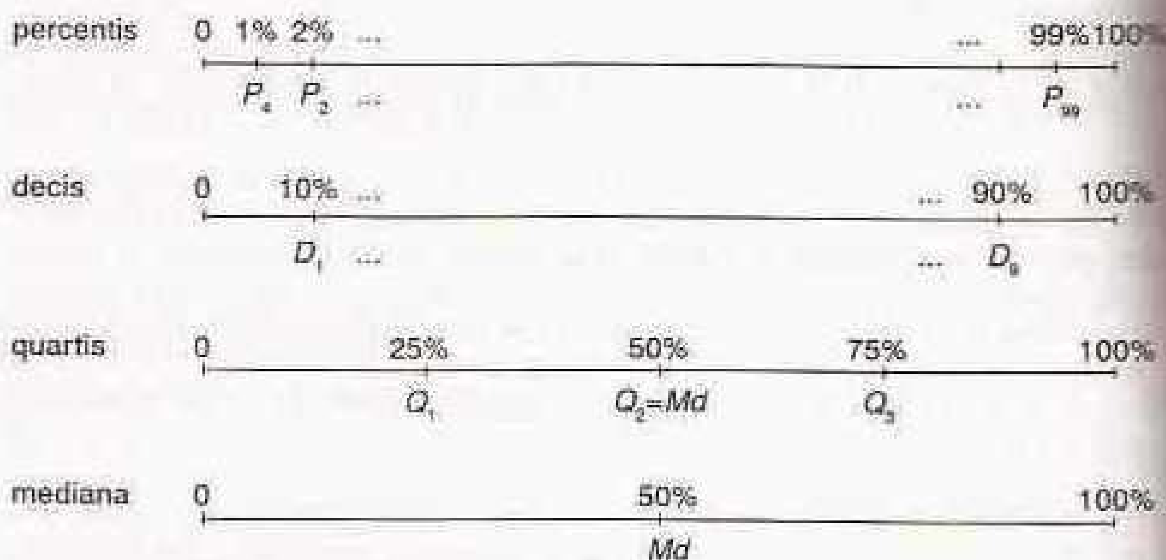
e) $P(\mu - 3{,}5\sigma \leq X \leq \mu + 3{,}5\sigma)$

20. Com base nas respostas obtidas no exercício anterior, pode-se concluir que: para qualquer distribuição Normal:

a) o intervalo compreendido entre o valor da média menos "um" desvio-padrão e o valor da média mais "um" desvio-padrão contém aproximadamente 68% das observações.

Elabore conclusões semelhantes, considerando os resultados obtidos nos outros itens do exercício 19.

21. Como você verá no capítulo 5, os *percentis* (Pi) são medidas estatísticas que dividem uma distribuição em 100 partes iguais. Da mesma forma os *decis* (Di) dividem a distribuição em dez partes iguais, enquanto os *quartis* (Qi) dividem a distribuição em quatro partes iguais, e a mediana (Md) divide a distribuição em duas partes iguais. Ou seja:



Determine o que se pede:

a) P_5 de uma $N\,(18; 64)$

b) P_{85} de uma $N\,(20; 100)$

c) D_1 de uma $N\,(30; 49)$

d) D_7 de uma $N\,(120; 81)$

e) Q_1 de uma $N\,(5; 9)$

f) Q_3 de uma $N\,(78; 121)$

g) Md de uma $N\,(30; 40)$

22. Usando a tabela da distribuição N (0, 1) determine Z_0 tal que:

a) $P\,(Z < Z_0) = 0{,}05$

b) $P\,(Z < Z_0) = 12\%$

c) $P\,(Z < Z_0) = 35\%$

d) $P\,(Z < Z_0) = 50\%$

e) $P\,(Z < Z_0) = 60\%$

f) $P\,(Z < Z_0) = 75\%$

g) $P\,(Z < Z_0) = 90\%$

h) $P\,(Z > Z_0) = 72\%$

i) $P\,(Z > Z_0) = 0{,}65$

j) $P\,(Z > Z_0) = 0{,}38$

l) $P\,(Z > Z_0) = 0{,}08$

Distribuição Exponencial

23. Uma lâmpada tem a duração de acordo com a densidade de probabilidade a seguir:

$$f(t) = \begin{cases} 0, & t < 0 \\ \dfrac{1}{1.000}\, e^{-\frac{t}{1.000}}, & t \geqslant 0 \end{cases}$$

Determinar:

a) a probabilidade de que uma lâmpada qualquer queime antes de 1.000 horas;

b) a probabilidade de que uma lâmpada qualquer queime depois de sua duração média;

c) qual é o desvio-padrão da distribuição.

24. Se as interrupções no suprimento de energia elétrica ocorrem segundo uma distribuição de Poisson com a média de uma interrupção por mês (quatro semanas), qual a probabilidade de que entre duas interrupções consecutivas haja um intervalo de:

a) menos de uma semana;

b) entre dez e doze semanas;

c) exatamente um mês;

d) mais de três semanas.

25. Prove que $f(t)$ é uma função densidade de probabilidade.

26. Prove que $\mu_{(t)} = \dfrac{1}{\lambda}$.

27. Prove que $\sigma^2_{(t)} = \dfrac{1}{\lambda^2}$.

4.4 DISTRIBUIÇÃO QUI-QUADRADO

Trata-se de um modelo de distribuição contínua muito importante para a teoria da inferência estatística.

Seja $x_1, x_2, \ldots, x_p$, "p" variáveis aleatórias independentes, normalmente distribuídas, com média zero e variância 1. Define-se variável aleatória com distribuição qui-quadrado, como:

$$\chi^2_p = x_1^2 + x_2^2 + \ldots + x_p^2$$

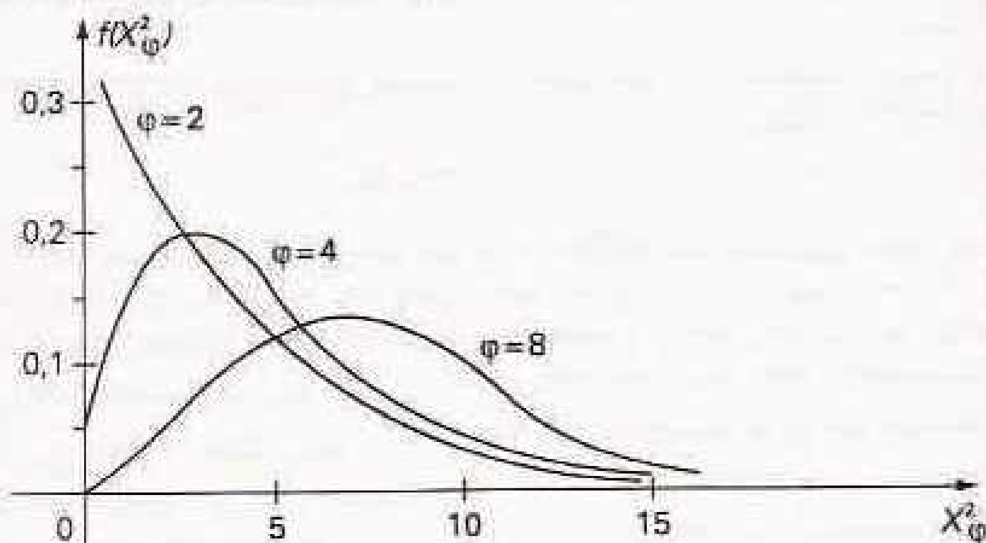
onde "p" é um parâmetro da função densidade denominado *grau de liberdade*. Normalmente utiliza-se a letra grega φ (lê-se fi) para se indicar o grau de liberdade.

Pode-se demonstrar que a média de uma distribuição qui-quadrado é igual ao grau de liberdade, e que a variância é igual ao dobro do número de graus de liberdade. Assim:

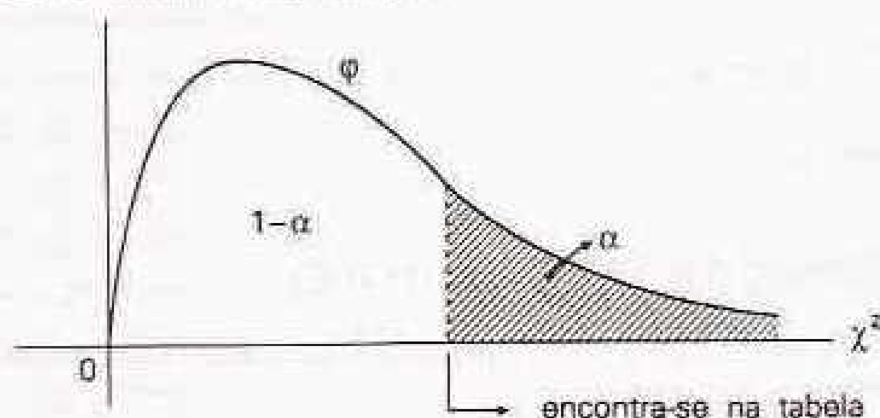
$$E[\chi^2_\varphi] = \mu(\chi^2_\varphi) = \varphi$$

$$Var[\chi^2_\varphi] = \sigma^2(\chi^2_\varphi) = 2\varphi$$

Conforme o número de graus de liberdade (valor do parâmetro), a curva que descreve a função densidade tem determinada forma. Eis algumas dessas curvas:



A distribuição qui-quadrado está tabelada. No anexo do livro você encontrará uma tabela que dá a abscissa da distribuição para diversas áreas (probabilidades) da cauda à direita. Assim:

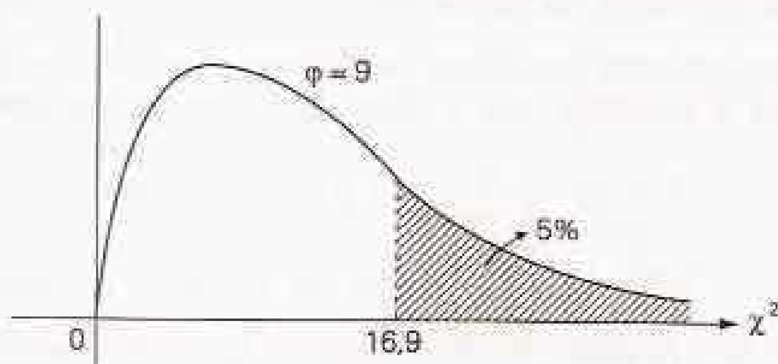


Na tabela tem-se:

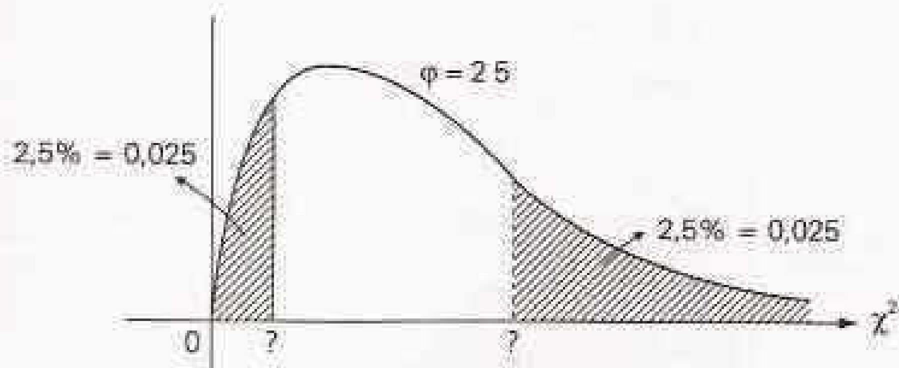


Exemplos: a) Admita parâmetro 9, ou seja, $\varphi = 9$ e $\alpha = 5\%$.

Entrando-se na 1ª coluna com $\varphi = 9$, e na 1ª linha com $\alpha = 0,05$, encontra-se na intersecção dessa linha e coluna o número 16,9. Graficamente tem-se:



b) Sendo $\varphi = 25$ e a seguinte distribuição:



O valor da abscissa à direita, chamado qui-quadrado superior, é obtido na tabela entrando-se na 1ª coluna com 25 e 1ª linha com 0,025. Assim: $\chi^2_{sup} = 40,6$.

O valor da abscissa à esquerda, chamado qui-quadrado inferior, é obtido na tabela entrando-se na 1ª coluna com 25 e 1ª linha com 0,975 = (1.000 – 0,025), já que a tabela considera áreas somente na cauda à direita, encontrando-se 13,1. Assim $\chi^2_{inf} = 13,1$.

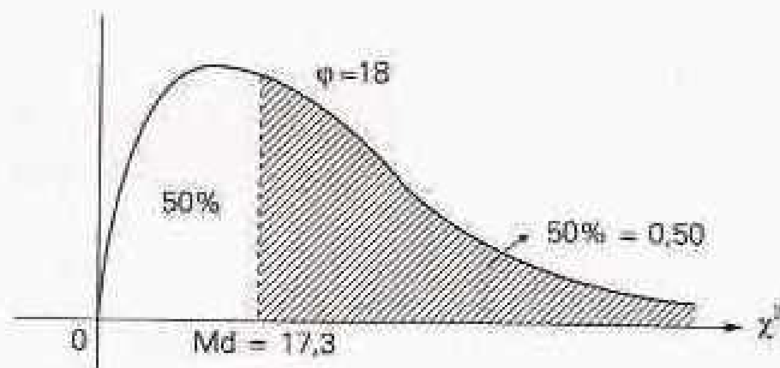
c) Considere uma distribuição qui-quadrado com parâmetro 18. Encontre a média variância, desvio-padrão, mediana, 1º quartil e 90º percentil.

Solução: *Média*: $\mu\,(\chi^2_{18}) = 18$

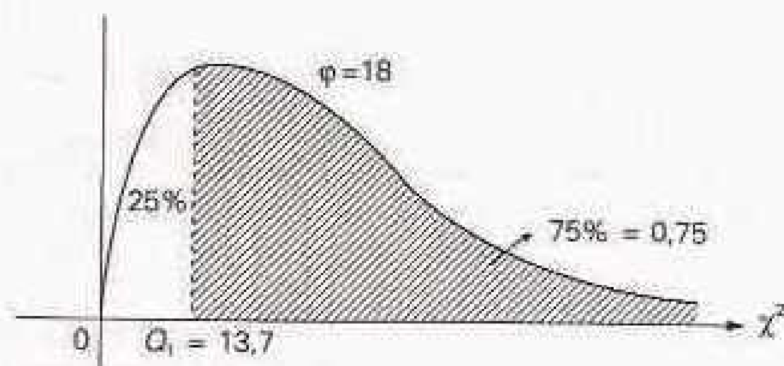
Variância: $\sigma^2\,(\chi_{18}) = 2(18) = 36$

Desvio-padrão: $\sigma\,(\chi^2_{18}) = \sqrt{36} = 6$

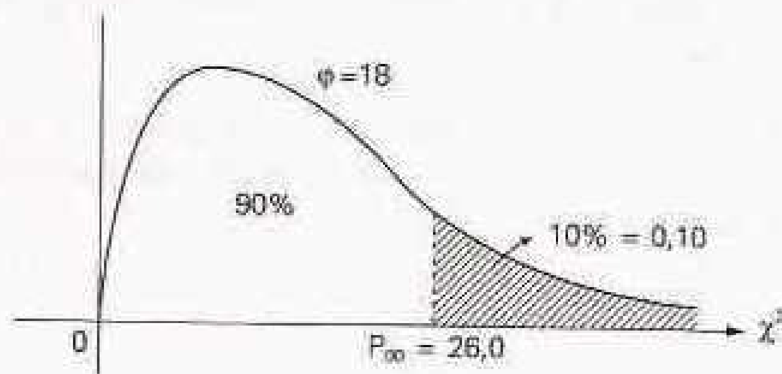
Mediana:



1º quartil:



90º percentil:



4.5 DISTRIBUIÇÃO t DE STUDENT

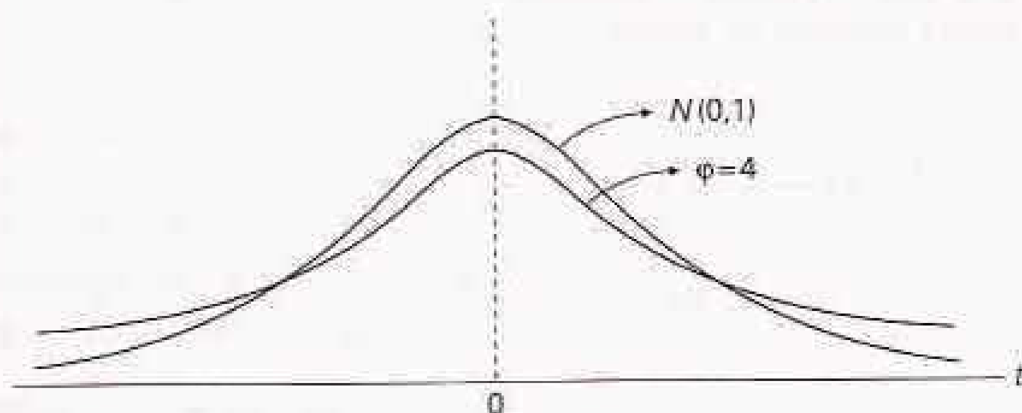
Trata-se de um modelo de distribuição contínua que se assemelha à distribuição normal padrão, $N(0,1)$. É utilizada para inferências estatísticas, particularmente, quando se tem amostras com tamanhos inferiores a 30 elementos.

A distribuição t também possui um parâmetro denominado grau de liberdade (φ). A média da distribuição é zero, e sua variância é dada por:

$$\text{Var}[t_\varphi] = \sigma^2(t_\varphi) = \frac{\varphi}{\varphi - 2} \quad (\varphi > 2)$$

A distribuição t é simétrica em relação à sua média.

Eis um exemplo da forma do gráfico da distribuição t de Student, quando $\varphi = 4$.



Observe que para valores de $\varphi < 30$ a distribuição "t" apresenta maior dispersão do que a $N(0,1)$, já que o desvio-padrão, nestes casos, é maior do que 1, que é o desvio-padrão da distribuição Normal Padrão. Por exemplo, para $\varphi = 4$, tem-se:

$$\sigma(t_4) = \sqrt{\frac{4}{4-2}} = 1{,}41$$

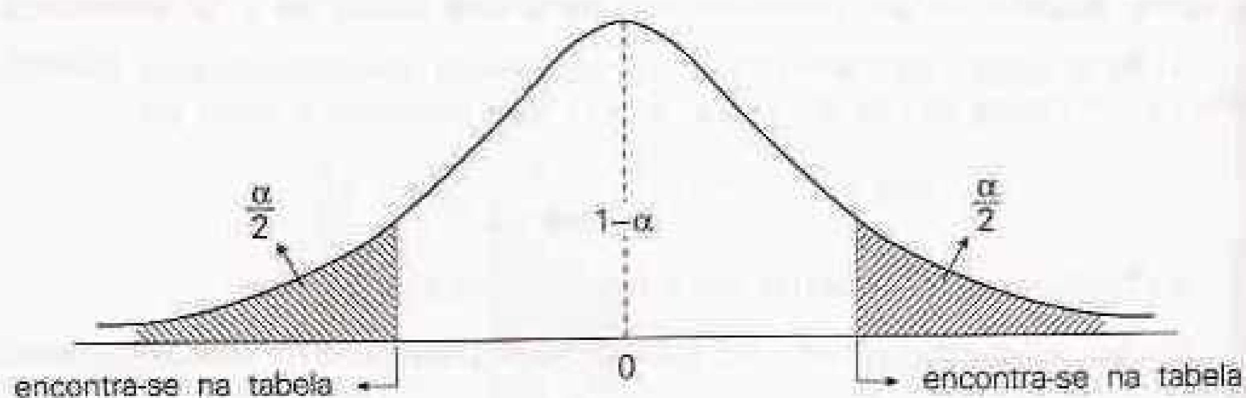
Se $\varphi = 35$ tem-se:

$$\sigma(t_{35}) = \sqrt{\frac{35}{35-2}} = 1{,}03$$

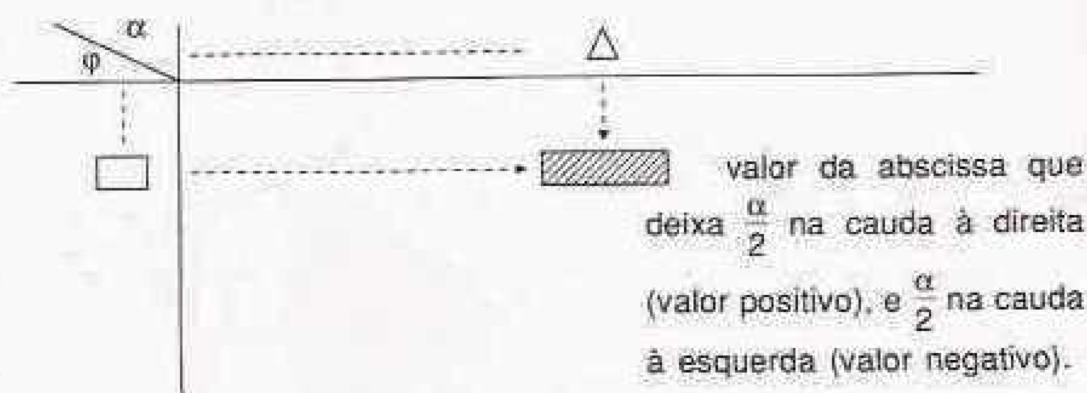
Se $\varphi = 60$ tem-se:

$$\sigma(t_{60}) = \sqrt{\frac{60}{60-2}} = 1{,}02$$

A distribuição t está tabelada. No anexo do livro você encontrará uma tabela que dá as abscissas da distribuição para diversas áreas (probabilidades) nas caudas. Trata-se de uma tabela bicaudal. Assim:



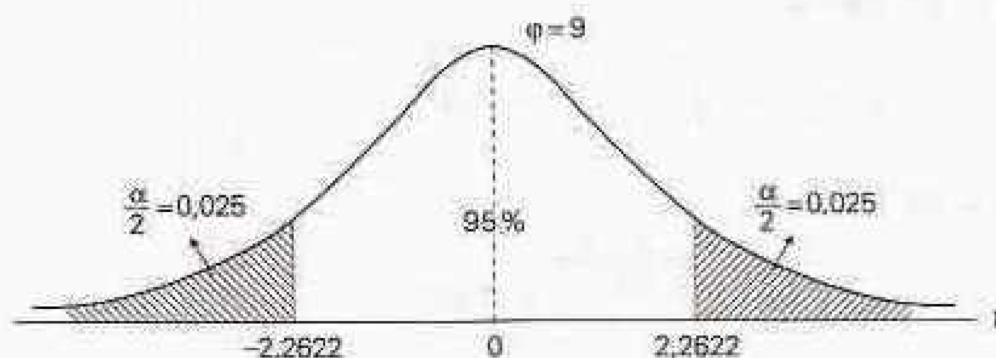
Na tabela procede-se assim:



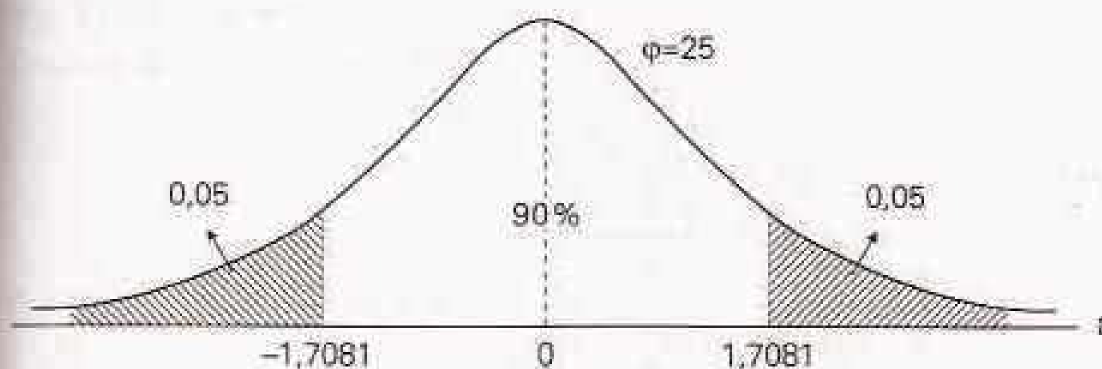
valor da abscissa que deixa $\frac{\alpha}{2}$ na cauda à direita (valor positivo), e $\frac{\alpha}{2}$ na cauda à esquerda (valor negativo).

Exemplos: a) Admita parâmetro 9, ou seja, $\varphi = 9$ e $\alpha = 5\%$.

Entrando-se na 1ª coluna com $\varphi = 9$, e na 1ª linha com $\alpha = 0,05$, encontra-se na intersecção dessa linha e coluna o número 2,2622. Graficamente tem-se:



b) Sendo $\varphi = 25$ e $\alpha = 10\%$. Tem-se:



Confira na tabela

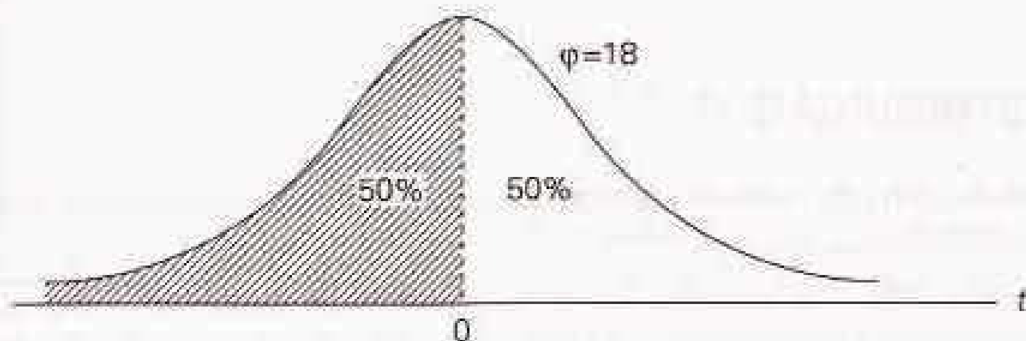
c) Considere uma distribuição t com parâmetro 18. Encontre a média, variância, desvio-padrão, mediana, 1º quartil e 95º percentil.

Solução: *Média*: $\mu(t_{18}) = 0$

Variância: $\sigma^2(t_{18}) = \dfrac{18}{18-2} = 1{,}13$

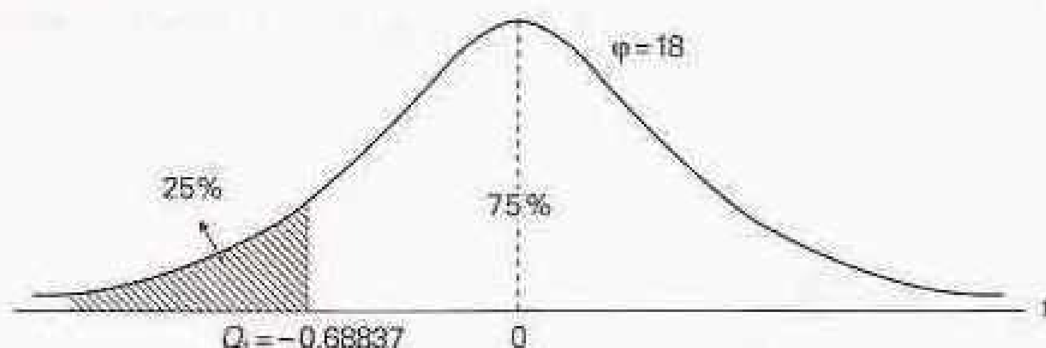
Desvio-padrão: $\sigma(t_{18}) = \sqrt{1{,}13} = 1{,}06$

Mediana:

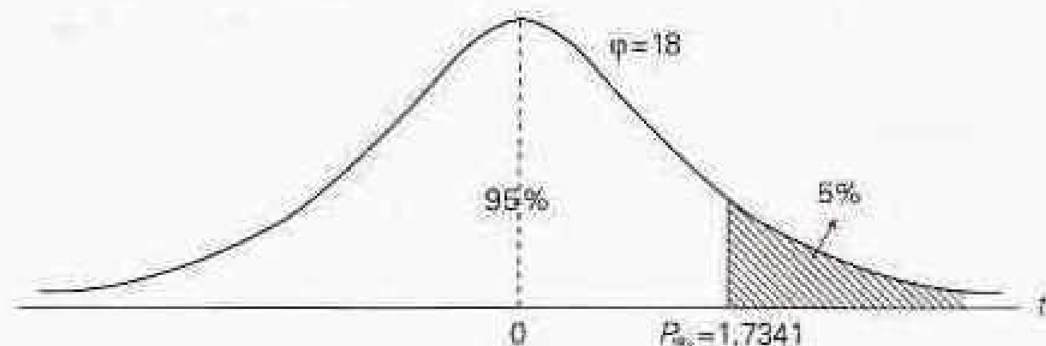


logo $Md(t_{18}) = 0$

1º quartil:



95º percentil:



4.6 DISTRIBUIÇÃO *F*

Trata-se de um modelo de distribuição contínua também útil para inferências estatísticas.

A distribuição *F* é a razão entre duas variáveis aleatórias independentes com distribuições qui-quadrado. Assim, uma distribuição *F* com "p" graus de liberdade no numerador, e "q" graus de liberdade no denominador é expressa por:

$$F(p, q) = \frac{\dfrac{\chi_p^2}{p}}{\dfrac{\chi_q^2}{q}} = \frac{\chi_p^2}{\chi_q^2} \cdot \frac{q}{p}$$

A distribuição F possui dois parâmetros: grau de liberdade do numerador e grau de liberdade do denominador, que são denominados, comumente, por φ_1 e φ_2 respectivamente.

Quanto à média, é dada por:

$$\boxed{\mu = \frac{\varphi_2}{\varphi_2 - 2}} \qquad (\varphi_2 > 2)$$

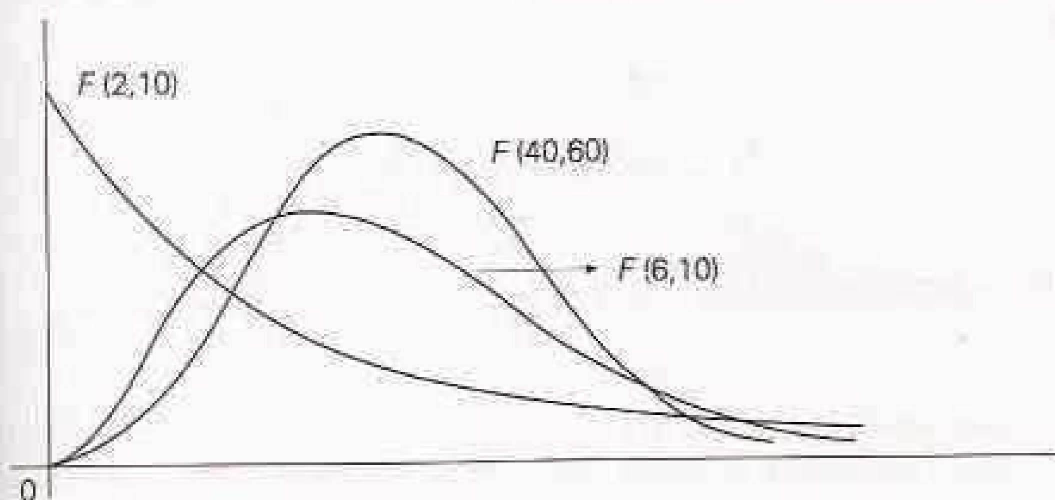
A variância é expressa por:

$$\boxed{\sigma^2 = \frac{2\varphi_2^2\,(\varphi_1 + \varphi_2 - 2)}{\varphi_1\,(\varphi_2 - 4)(\varphi_2 - 2)^2}} \qquad (\varphi_2 > 4)$$

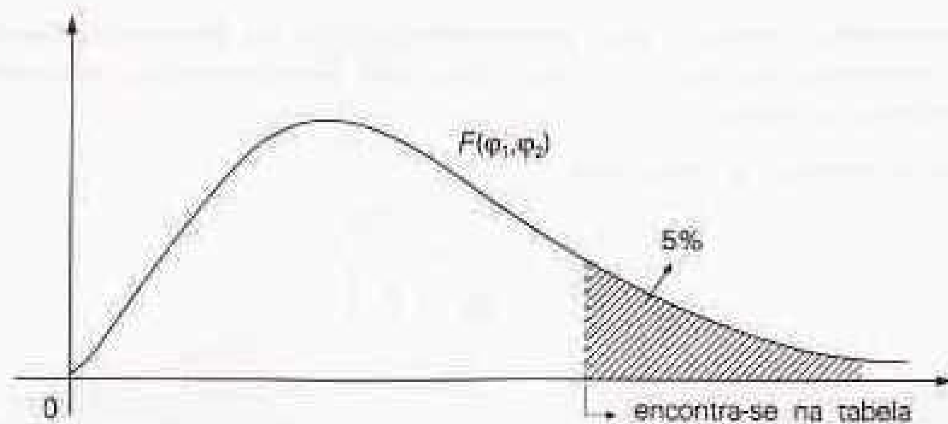
E a moda:

$$\boxed{M_0 = \left(\frac{\varphi_1 - 2}{\varphi_1}\right)\left(\frac{\varphi_2}{\varphi_2 + 2}\right)} \qquad (\varphi_1 > 2)$$

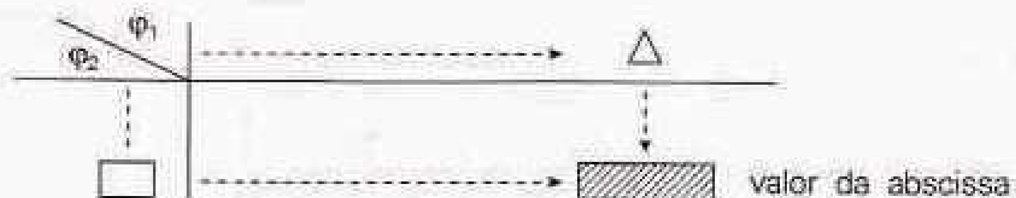
Eis algumas formas de gráficos da distribuição F para determinados graus de liberdade:



A distribuição F está tabelada. No anexo do livro você encontrará uma tabela que dá as abscissas que deixam 5% na cauda à direita, dados os parâmetros φ_1 e φ_2. Assim:



Na tabela procede-se assim:



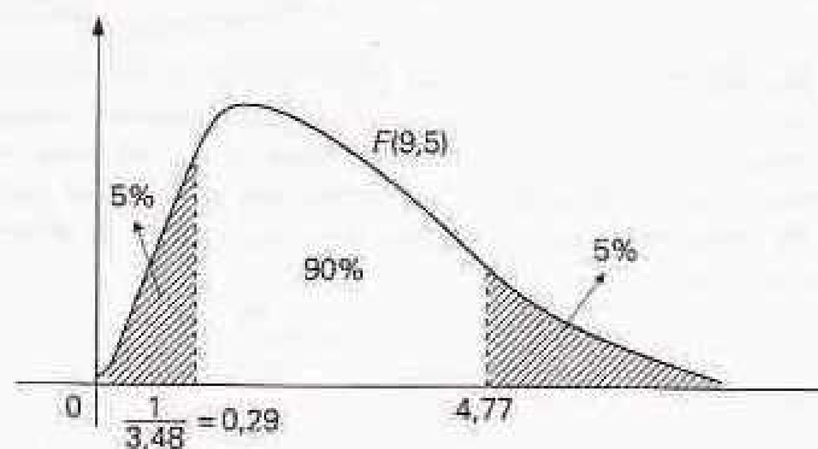
Existem tabelas que expressam abscissas para outros níveis de probabi-lidade na cauda à direita.

Para se encontrar o valor da abscissa $F_{1-\alpha}(\varphi_1, \varphi_2)$ utiliza-se a seguinte fórmula:

$$F_{1-\alpha}(\varphi_1, \varphi_2) = \frac{1}{F\alpha(\varphi_2, \varphi_1)}$$

Exemplo: Sendo $\varphi_1 = 9$, $\varphi_2 = 5$ e $\alpha = 5\%$.

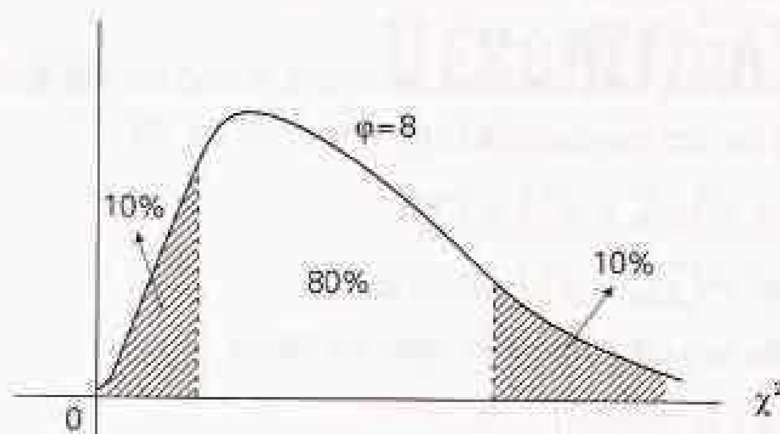
logo:



EXERCÍCIOS – SÉRIE II – CAPÍTULO 4

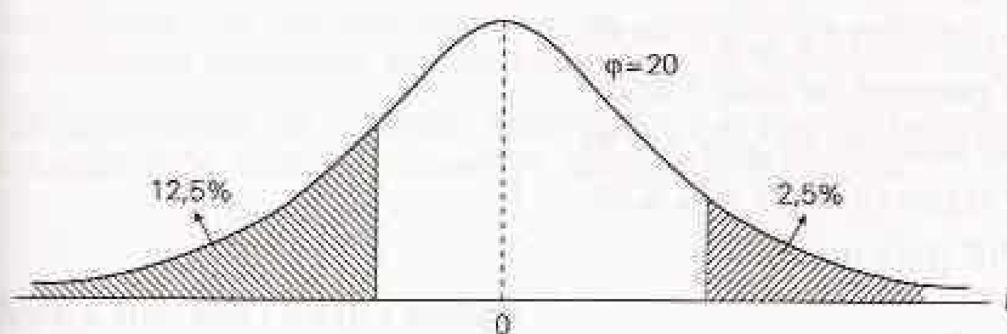
1. Considere uma distribuição qui-quadrado, com 23 graus de liberdade. Determine a média, variância, desvio-padrão, mediana e 3º quartil.

2. Determine os valores do χ^2_{sup} e χ^2_{inf}

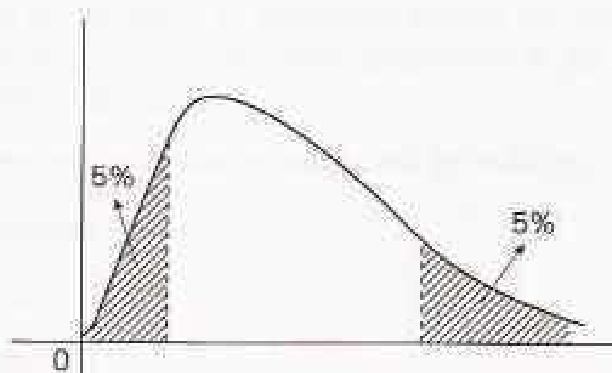


3. Considere uma distribuição *t* com parâmetro 23. Determine a média, variância, desvio-padrão, 1º quartil, o 5º percentil e moda.

4. Consulte a tabela para descobrir os valores das abscissas.



5. Admite uma distribuição *F* com $\varphi_1 = 8$ e $\varphi_2 = 10$. Determine a média, variância, desvio-padrão e moda bem como as abscissas para:



6. Determine o que se pede:

a) 1º quartil de uma distribuição $N\,(100; 49)$.

b) Z_0, tal que: $P\,(Z > Z_0) = 65\%$.

c) Z_0, tal que: $P\,(Z < Z_0) = 80\%$.

d) $P\,(-1{,}57 \leq Z \leq 2{,}42)$.

e) 90º percentil de uma distribuição N (R\$ 2.000,00; R\$ 2.225,00).

f) Mediana de um qui-quadrado com parâmetro 30.

g) X^2 tal que: $P\left(\chi^2_{15} < \chi^2\right) = 10\%$.

h) X^2 tal que: $P\left(\chi^2_{20} > \chi^2\right) = 0{,}25\%$.

i) 9º decil de um qui-quadrado com variância 50.

j) $P\left(13{,}8 \leq \chi^2_{26} \leq 38{,}9\right)$.

l) 3º quartil de uma "t" de Student com $\varphi = 5$.

m) $P\,(t_8 > 2{,}3060)$.

n) $P\,(t_{14} < -2{,}9768)$.

o) $P\,(-1{,}1816 \leq t_{22} \leq 3{,}1188)$.

p) 95º percentil de uma "t" com parâmetro 27.

q) $P\,(-0{,}68276 \leq t_{30} \leq 2{,}7500)$.

r) 5º percentil de uma $F\,(8,7)$.

s) 95º percentil de uma $F\,(7,8)$.

t) $P\,(0{,}00418 \leq F\,(1,8) \leq 5{,}32)$.

u) $P\,(F\,(6,4) < 0{,}22075)$.

5

Estatística Descritiva

5.1 INTRODUÇÃO

Como o próprio nome sugere, estatística descritiva se constitui num conjunto de técnicas que objetivam descrever, analisar e interpretar os dados numéricos de uma população ou amostra.

Neste capítulo serão apresentados exemplos de tabelas e gráficos que podem representar, de maneira sintética, as informações sobre o comportamento de variáveis numéricas levantadas através de processos de pesquisa.

Serão destacadas as medidas estatísticas que podem representar uma amostra: medidas de posição, dispersão, assimetria e curtose.

5.2 TABELAS ESTATÍSTICAS

Existem regulamentações para construção de tabelas estatísticas, todavia, em função dos objetivos práticos deste livro, não serão adotadas normas rígidas para a elaboração de tabelas. Tendo-se claro que a tabela deverá ser uma forma objetiva de se demonstrar o comportamento de variáveis, o que se deve buscar são representações simples que possibilitem ao leitor a compreensão do fenômeno sem muito esforço.

Uma tabela deve apresentar a seguinte estrutura;

- cabeçalho;
- corpo;
- rodapé.

O cabeçalho deve conter o suficiente para que sejam respondidas as questões:

- O que está representado?
- Onde ocorreu?
- Quando ocorreu?

O corpo da tabela é representado por colunas e subcolunas dentro das quais serão registrados os dados numéricos e informações.

O rodapé é reservado para observações pertinentes à tabela, bem como para o registro e identificação da fonte dos dados.

Conforme critério de agrupamento as tabelas podem representar diversas séries. Assim:

1. *Série Cronológica, Temporal, Evolutiva ou Histórica.*

É a série estatística em que os dados são observados segundo a época de ocorrência. Exemplo:

Vendas da Companhia Alfa – 1970-1977

Ano	*Vendas (em $ 1.000,00)*
1970	2.181
1971	3.948
1972	5.642
1973	7.550
1974	10.009
1975	11.728
1976	18.873
1977	29.076

Fonte: Departamento e Marketing da Companhia.

2. *Série Geográfica ou de Localização.*

É a série estatística em que os dados são observados segundo a localidade de ocorrência. Exemplo:

INAMPS – Empresas fiscalizadas em 1973

Regiões	*Empresas fiscalizadas*
Norte	7.495
Nordeste	107.783
Sudeste	281.207
Sul	53.661
Centro-Oeste	15.776

Fonte: Mensário Estatístico 259/260.

3. *Série Específica:*

É a série estatística em que os dados são agrupados segundo a modalidade de ocorrência. Exemplo:

Matrícula no Ensino de Terceiro Grau
Brasil – 1975 (ciclo básico)

Áreas de ensino	*Matrículas*
Ciências Biológicas	32.109
Ciências Exatas e Tecnologia	65.949
Ciências Agrárias	2.419
Ciências Humanas	148.842
Letras	9.883
Artes	7.464
Duas ou mais áreas	16.323

Fonte: Serviço de Estatística da Educação e Cultura.

4. *Distribuição de Freqüências.*

É a série estatística em que os dados são agrupados com suas respectivas freqüências absolutas. Exemplos:

a) *Número de Acidentes por Dia na Rodovia X em Janeiro de 1977.*

Número de acidentes por dia	*Número de dias*
0	10
1	7
2	4
3	5
4	3
5	2

Fonte: DNER.

b) *Altura dos Alunos da Classe em Março de 1977.*

Alturas (m)	*Número de alunos*
1,50 ⊢ 1,60	5
1,60 ⊢ 1,70	15
1,70 ⊢ 1,80	17
1,80 ⊢ 1,90	3

Fonte: Secretaria da escola.

5.3 Gráficos

A representação gráfica das séries estatísticas tem por finalidade representar os resultados obtidos, permitindo chegar-se a conclusões sobre a evolução do fenômeno ou sobre como se relacionam os valores da série. Não há uma única maneira de representar graficamente uma série estatística. A escolha do gráfico mais apropriado ficará a critério do analista. Contudo, os elementos simplicidade, clareza e veracidade devem ser considerados quando da elaboração de um gráfico.

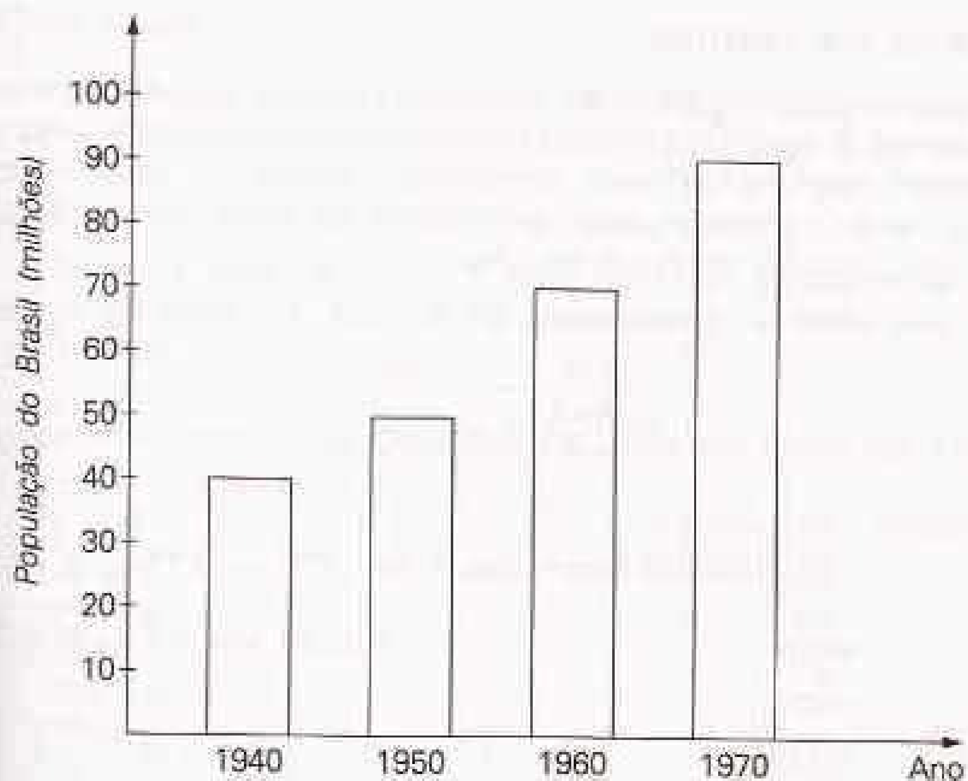
Eis os principais tipos de gráficos.

1. *Gráfico em Colunas*

População Brasileira – 1940-1970

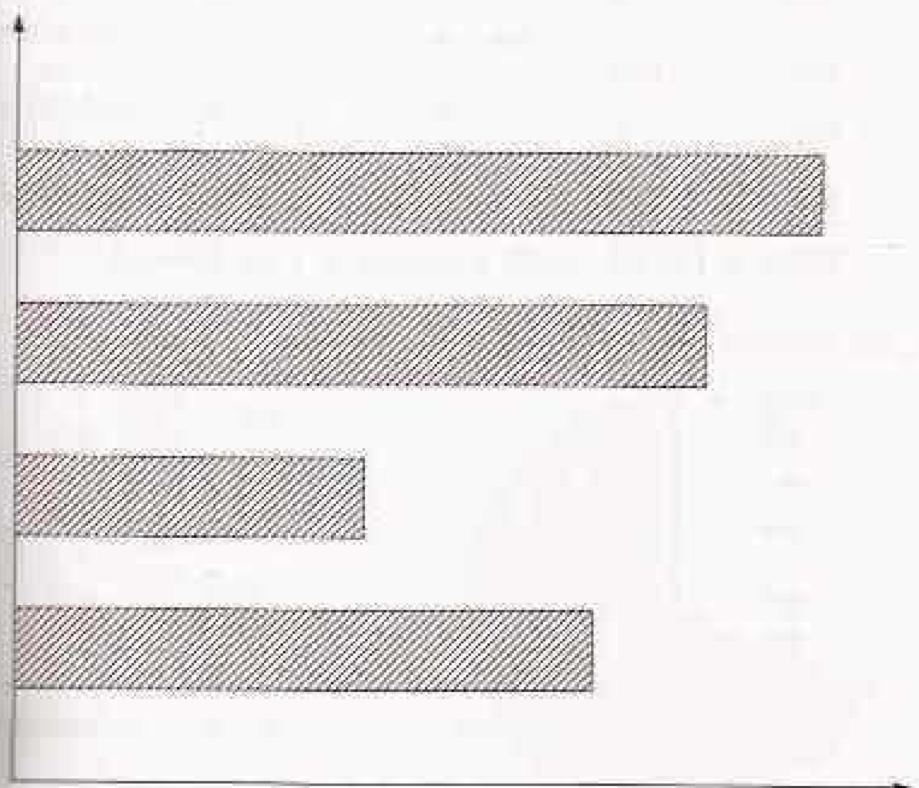
Ano	*População*
1940	41.236.315
1950	51.944.397
1960	70.119.071
1970	93.139.037

Fonte: Anuário Estatístico – 1974.



2. *Gráfico em Barras*

É semelhante ao gráfico em colunas, porém os retângulos são dispostos horizontalmente. Eis uma configuração:



3. *Gráficos em Setores*

É a representação gráfica de uma série estatística, em um círculo, por meio de setores. É utilizado principalmente quando se pretende comparar cada valor da série com o total. Para construí-lo, divide-se o círculo em setores, cujas áreas serão proporcionadas aos valores da série. Essa divisão poderá ser obtida pela solução da regra de três.

total ______	360°
parte ______	x°

Exemplo:

Receita do Município X de 1975 a 1977

Anos	*Receita (em Cr$ 1.000.000,00)*
1975	90
1976	120
1977	150
Total	360

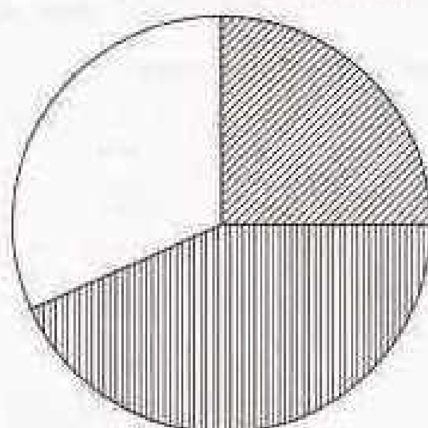
Fonte: Departamento da Fazenda, Município X.

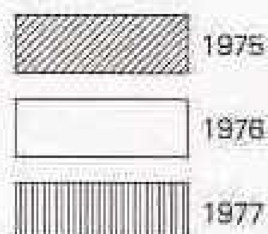
para 1975: $360 - 360^\circ$
$90 - x^\circ$ $\quad x^\circ = 90^\circ$

para 1976: $360 - 360^\circ$
$120 - x^\circ$ $\quad x^\circ = 120^\circ$

para 1977: $360 - 360^\circ$
$150 - x^\circ$ $\quad x^\circ = 150^\circ$

Receita do Município X





4. *Gráfico Polar*

É a representação de uma série por meio de um polígono. Geralmente presta-se para apresentação de séries temporais. Para construí-lo, divide-se uma circunferência em tantos arcos iguais quantos forem os dados a representar. Pelos pontos de divisas traçam-se raios.

Em cada raio é representado um valor da série, marcando-se um ponto cuja distância ao centro é diretamente proporcional a esse valor. A seguir unem-se os pontos. Exemplo:

Movimento Mensal de Compras da Empresa Delta em 1972

Meses	*Valores (Cr$ 1.000,00)*
Janeiro	12
Fevereiro	13
Março	14
Abril	12
Maio	15
Junho	19
Julho	17
Agosto	18
Setembro	14
Outubro	16
Novembro	12
Dezembro	18

Fonte: Departamento de compras da empresa Delta.

Veja o gráfico na próxima página.

5. *Gráficos em Curvas*

Vendas da Companhia Beta – 1971 a 1977

Ano	*Vendas (Cr$ 1.000,00)*
1971	230
1972	260
1973	380
1974	300
1975	350
1976	400
1977	450

Fonte: Departamento de Marketing da Companhia.

Veja o gráfico na próxima página.

Gráfico Polar

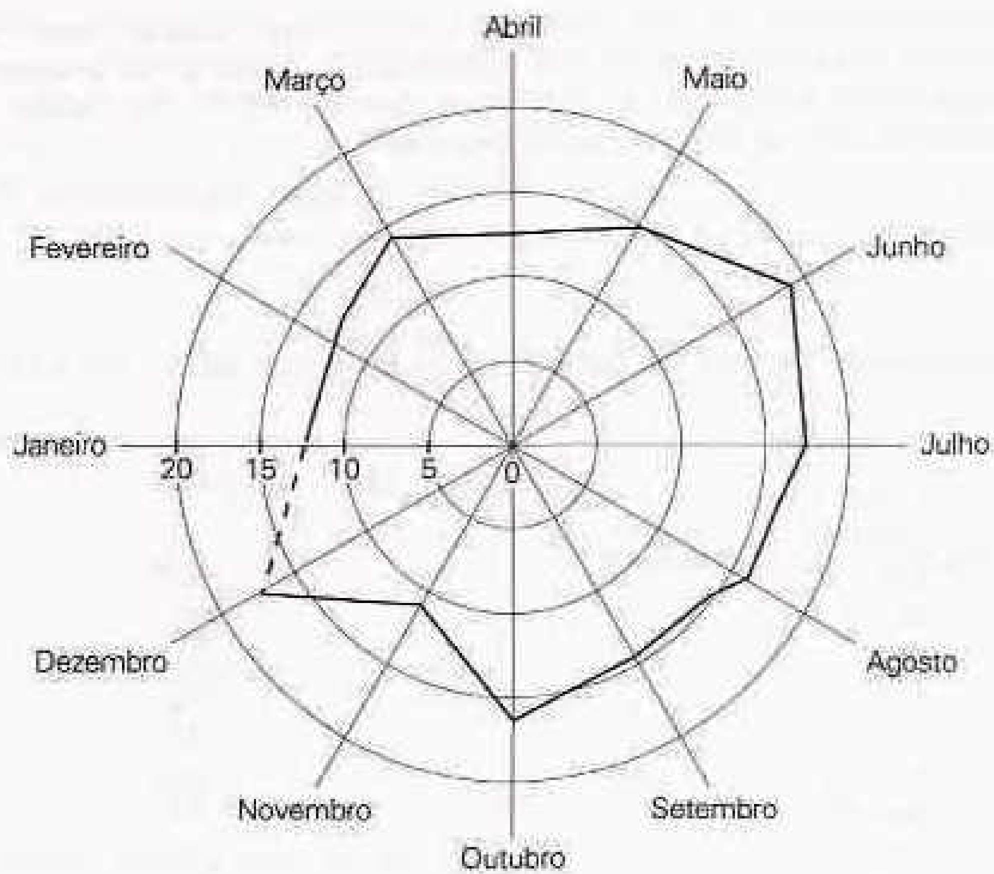
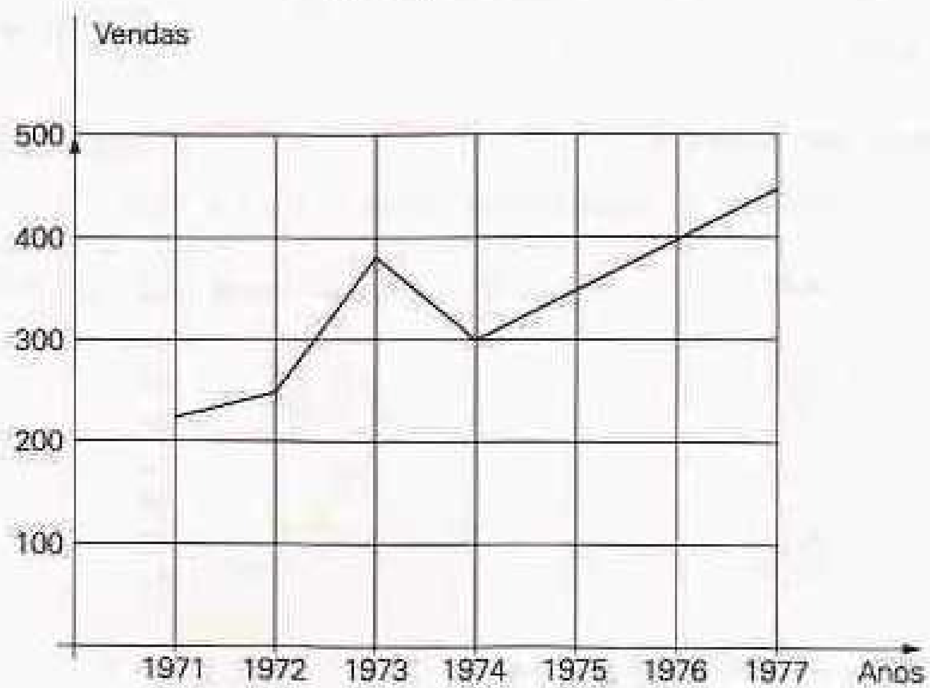


Gráfico em Curvas



De uso mais restrito, têm-se os cartogramas (representação por intermédio de carta geográfica), os pictogramas (apresentação de uma série estatística por meio de símbolos representativos do fenômeno) e os estereogramas (representação gráfica de uma série estatística por meio de corpos sólidos geométricos).

EXERCÍCIOS – SÉRIE I – CAPÍTULO 5

1. Montar uma série cronológica para representar os valores das exportações de açúcar, fornecidas pelo Instituto do Açúcar e do Álcool, nos anos de 1965 a 1971 em milhares de dólares: 60.193 – 80.114 – 812.826 – 106.879 – 112.064 – 126.740 – 149.548.

2. Idealizar uma série geográfica para representar o seguinte fato: população da região Norte do Brasil em 1970, sabendo-se que em Rondônia, Acre, Amazonas, Roraima, Pará e Amapá, temos, respectivamente: 116.620 – 218.006 – 960.934 – 41.638 – 2.197.072 e 116.480 habitantes, segundo dados da Fundação IBGE.

3. Fazer uma tabela estatística para representar o movimento religioso de certo município no período 1975-1977, que apresentou os seguintes dados: em 1975, houve 56.738 habitantes batizados (dos quais 26.914 do sexo feminino), 15.884 casamentos e 13.678 extremas-unções. Em 1976, houve 33.915 batizados do sexo masculino e 29.568 do sexo feminino; os casamentos foram em número de 17.032 e as extremas-unções, 14.328. Em 1977, em um total de 71.232, 34.127 eram do sexo masculino; as extremas-unções foram 16.107 e os casamentos 16.774.

4. A tabela a seguir mostra as áreas, em milhões de km^2, dos oceanos. Representar graficamente os dados, usando: a) um gráfico de colunas; b) um gráfico de setores.

Oceano	*Antártico*	*Ártico*	*Atlântico*	*Índico*	*Pacífico*
Área (milhões km^2)	36,8	23,2	199,4	137,9	342,7

5. Representar em um gráfico polar os dados:

Meses	J	F	M	A	M	J	J	A	S	O	N	D
Temperatura (°C)	28	29	27	24	20	19	18	21	22	24	28	30

6. Construir um gráfico em barras que represente a série:

INAMPS – Benefícios Concedidos – Brasil – 1973

Espécie	*Quantidade*
Auxílio-natalidade	901.000
Auxílio-doença	467.000
Auxílio-funeral	88.000
Aposentadoria por Invalidez	40.000
Aposentadoria por Tempo de Serviço	39.000
Abono Permanente em Serviço	30.000
Pensão por Morte	73.000
Outras Espécies	44.000

Fonte: Mensário Estatístico do INAMPS.

7. Usando um gráfico em curva, representar a tabela a seguir:

Índices dos Preços Recebidos pelos Agricultores no Brasil em 1976 (1986 = 100)

Meses	*Índices*		
	Lavoura	*Produtos Animais*	*Agropecuário*
Janeiro	1.304	884	1.044
Fevereiro	1.418	891	1.092
Março	1.494	916	1.136
Abril	1.580	943	1.186
Maio	1.715	964	1.250
Junho	1.816	960	1.287
Julho	1.929	972	1.337
Agosto	2.013	1.015	1.396
Setembro	2.113	1.066	1.473
Outubro	2.197	1.097	1.517
Novembro	2.290	1.119	1.566
Dezembro	2.358	1.144	1.607

Fonte: IBGE.

5.4 DISTRIBUIÇÃO DE FREQÜÊNCIAS

Por constituir-se no tipo de tabela mais importante para a Estatística Descritiva, será apresentado um estudo completo das distribuições de freqüências. A seguir são descritos os procedimentos usuais na construção dessas tabelas. Eis alguns conceitos fundamentais:

5.4.1 População

É um conjunto de indivíduos ou objetos que apresentam pelo menos uma característica em comum. A população pode ser finita ou infinita. Na prática, quando uma população é finita, com um número grande de elementos, considera-se como população infinita.

5.4.2 Amostra

Considerando-se a impossibilidade, na maioria das vezes, do tratamento de todos os elementos da população, retira-se uma amostra. Para os propósitos dessa apresentação, admite-se que uma amostra já tenha sido escolhida de conformidade com alguma técnica de amostragem.

5.4.3 Variável Discreta e Variável Contínua

A variável é discreta quando assume valores em pontos da reta real. Exemplo: número de erros em um livro: 0, 1, 2, 3, 4, 5, ...

Por outro lado, quando a variável pode assumir teoricamente qualquer valor em certo intervalo da reta real, ela será uma variável contínua. Exemplo: peso de alunos, pois, teoricamente, um indivíduo poderá ter 50,5 kg; 50,572 kg; 50,585 kg; ...

5.4.4 Representação da Amostra

Como se observou anteriormente, a Estatística tem como objetivo encontrar leis de comportamento para todo o conjunto, por meio da sintetização dos dados numéricos, sob a forma de tabelas, gráficos e medidas. A seguir são apresentados os procedimentos para a representação das distribuições de freqüências.

1. *Dados brutos*

O conjunto dos dados numéricos obtidos após a crítica dos valores coletados constitui-se nos dados brutos. Assim: 24 – 23 – 22 – 28 – 35 – 21 – 23 – 33 – 34 – 24 – 21 – 25 – 36 – 26 – 22 – 30 – 32 – 25 – 26 – 33 – 34 – 21 – 31 – 25 – 31 – 26 – 25 – 35 – 33 – 31 são exemplos de dados brutos.

2. *Rol*

É o arranjo dos dados brutos em ordem crescente ou decrescente. Assim: 21 – 21 – 21 – 22 – 22 – 23 – 23 – 24 – 25 – 25 – 25 – 25 – 26 – 26 – 26 – 28 – 30 – 31 – 31 – 31 – 32 – 33 – 33 – 33 – 34 – 34 – 34 – 35 – 35 – 36 constituem o rol.

3. *Amplitude total ou "range" (R)*

É a diferença entre o maior e o menor valor observados. No exemplo, $R = 36 - 21 = 15$.

4. *Freqüência absoluta (F_i)*

É o número de vezes que o elemento aparece na amostra, ou o número de elementos pertencentes a uma classe. No exemplo, $F_{(21)} = 3$.

5. *Distribuição de freqüência*

É o arranjo dos valores e suas respectivas freqüências. Assim, a distribuição de freqüências para o exemplo será:

X_i	F_i
21	3
22	2
23	2
24	1
25	4
26	3
28	1
30	1
31	3
32	1
33	3
34	3
35	2
36	1
Σ	30

Exemplo de distribuição de freqüências de uma variável discreta (tipo A)

Observe: 1. X representa a variável.

2. $\Sigma F_i = n$.

3. n = tamanho da amostra.

Exemplo de distribuição de freqüências para variável contínua:

Seja X_i peso de 100 indivíduos:

Classe	F_i
45 ⊢ 55	15
55 ⊢ 65	30
65 ⊢ 75	35
75 ⊢ 95	15
85 ⊢ 95	5
Σ	100

Exemplo da distribuição de freqüências de uma variável contínua (tipo B)

6. *Número de classes (K)*

Não há uma fórmula exata para o cálculo do número de classes.

a) $K = 5$ para $n \leqslant 25$ e $K \cong \sqrt{n}$, para $n > 25$.

b) Fórmula de Sturges $K \cong 1 + 3{,}22 \log n$, em que n = tamanho da amostra.

Exemplo: Seja $n = 49$, então:

$$K = \sqrt{49} = 7 \text{ ou } K \cong 1 + 3{,}22 \log 49 \cong 7.$$

7. *Amplitude das classes (h)*

$$h \cong R \div K$$

Assim como no caso do número de classes (K), a amplitude das classes (h) deve ser aproximada para o maior inteiro. Assim, se $K \cong 6{,}4$, usa-se $K = 7$ ou $h \cong 1{,}7$, usa-se $h = 2$.

8. *Limites das classes*

Existem diversas maneiras de expressar os limites das classes. Eis algumas:

a) 10 ⊢⊣ 12: compreende todos os valores entre 10 e 12;

b) 10 ⊢ 12: compreende todos os valores de 10 a 12, excluindo o 12;

c) limite aparente 10 – 12; limite real 9,5 – 11,5;

d) 10 ⊣ 12 compreende todos os valores, excluindo o 10.

Usaremos a forma expressa no exemplo b.

9. *Pontos médios das classes (x_i)*

É a média aritmética entre o limite superior e o limite inferior da classe. Assim, se a classe for 10 |- 12, teremos:

$$x_i = \frac{10 + 12}{2} = 11,$$ como ponto médio da classe.

10. *Freqüência absoluta acumulada (F_{ac})*

É a soma das freqüências do valores inferiores ou iguais ao valor dado. **Exemplo:**

x_i	F_i	F_{ac}
0	3	3
1	5	8
2	2	10
Σ	10	

11. *Freqüência relativa (f_i).*

A freqüência relativa de um valor é dada por $f_i = \frac{F_i}{n}$, ou seja, é a porcentagem daquele valor na amostra. Exemplo:

X_i	F_i	f_i
1	5	5/14
2	7	1/2
3	2	1/7
Σ	14	1

Note: $\Sigma f_i = 1$. Assim 1/2 = 50% dos elementos correspondem ao 2.

12. *Histograma*

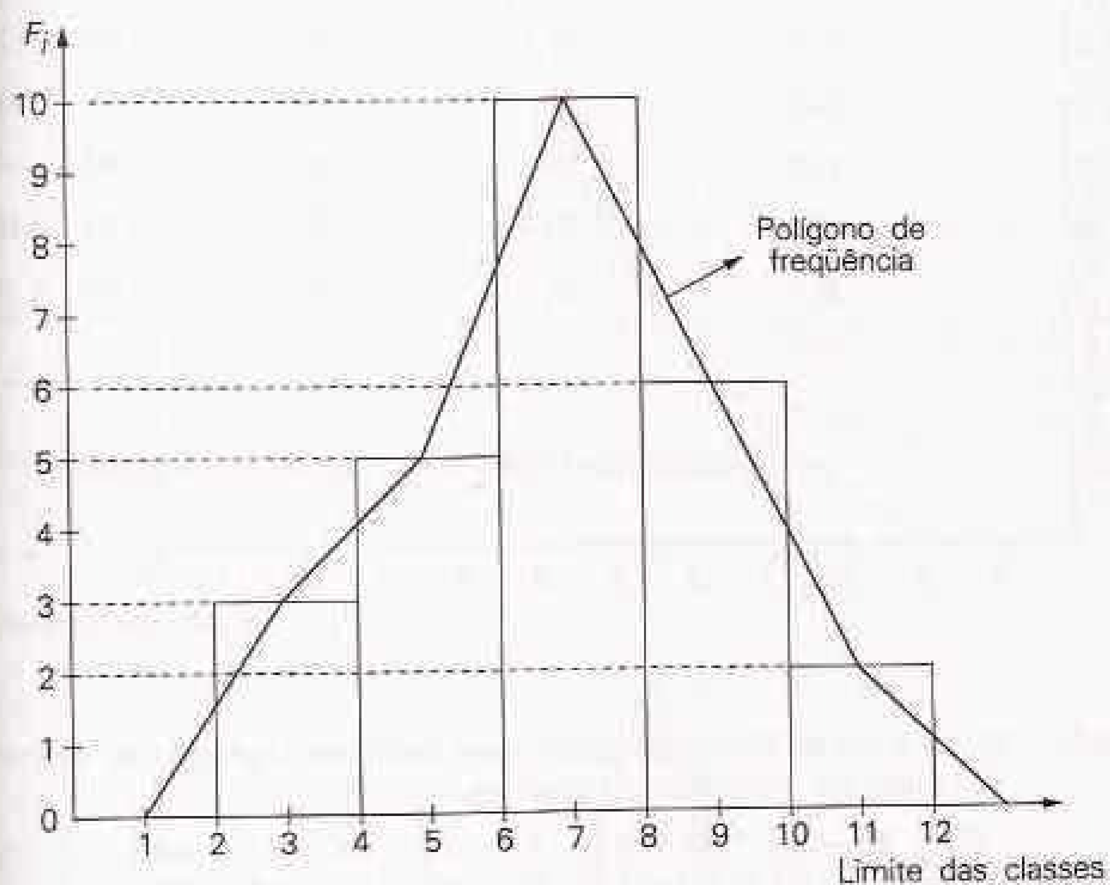
É a representação gráfica de uma distribuição de freqüência por meio de retângulos justapostos (veja exemplo a seguir).

13. *Polígono de freqüências*

É a representação gráfica de uma distribuição por meio de um polígono. Exemplo (histograma e polígono de freqüências):

Idade dos Alunos da Escola

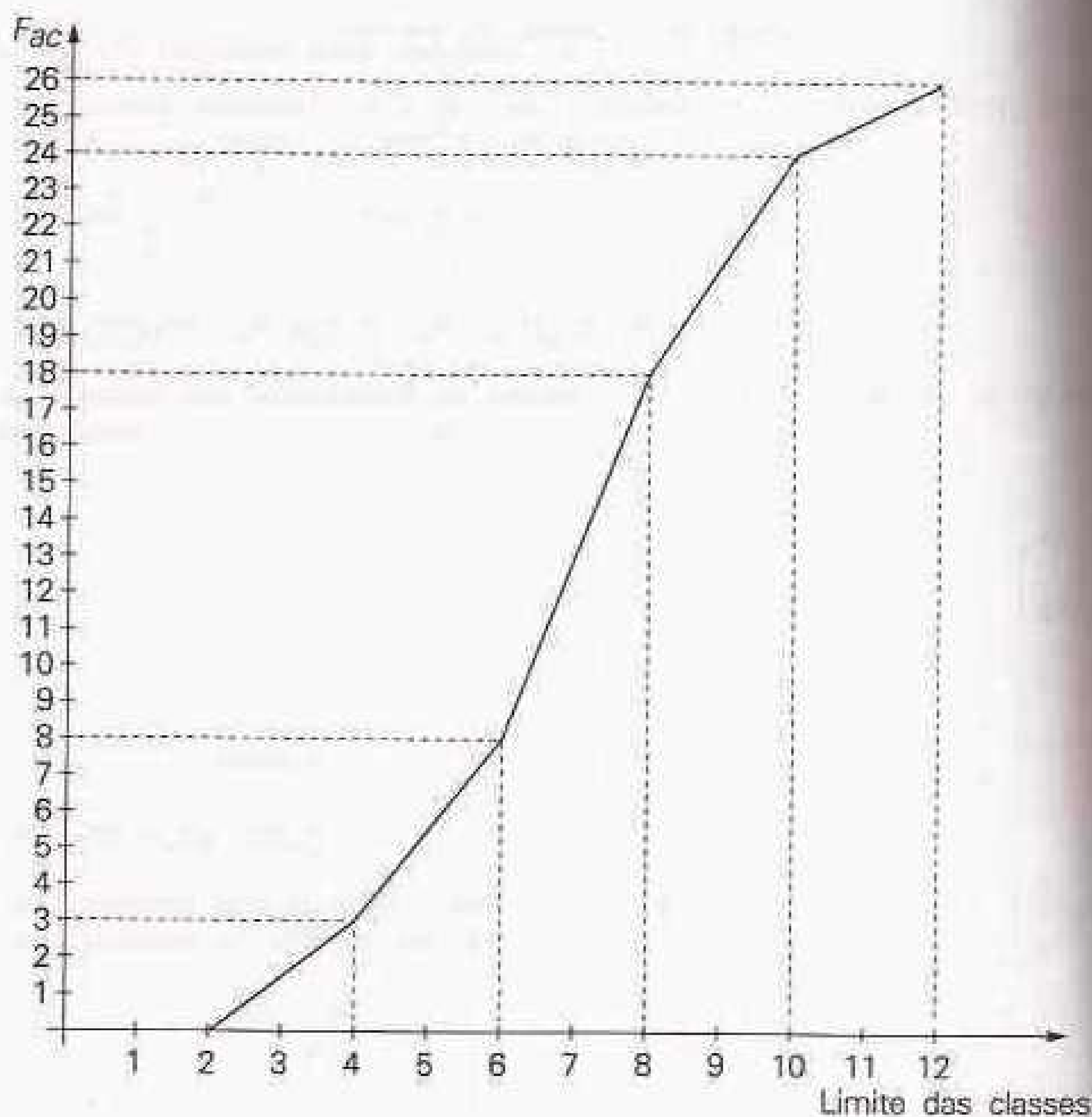
Idade	F_i
2 ⊢ 4	3
4 ⊢ 6	5
6 ⊢ 8	10
8 ⊢ 10	6
10 ⊢ 12	2
Σ	26



14. *Polígono de freqüência acumulada*

Exemplo:

Classes	2 ⊢ 4	4 ⊢ 6	6 ⊢ 8	8 ⊢ 10	10 ⊢ 12
F_i	3	5	10	6	2
F_{ac}	3	8	18	24	26



Exemplo: Dado o rol de 50 notas (dadas em créditos), agrupar os elementos em classe e construir os gráficos:

33 – 35 – 35 – 39 – 41 – 41 – 42 – 45 – 47 – 48
50 – 52 – 53 – 54 – 55 – 55 – 57 – 59 – 60 – 60
61 – 64 – 65 – 65 – 65 – 66 – 66 – 66 – 67 – 68
69 – 71 – 73 – 73 – 74 – 74 – 76 – 77 – 77 – 78
80 – 81 – 84 – 85 – 85 – 88 – 89 – 91 – 94 – 97

Solução: Amplitude total (R)

$$R = 97 - 33 = 64$$

Número de classes (K):

$$K \cong 1 + 3{,}22 \log 50 \cong 1 + 3{,}22(1{,}7) \cong 7$$

Amplitude de classe (h):

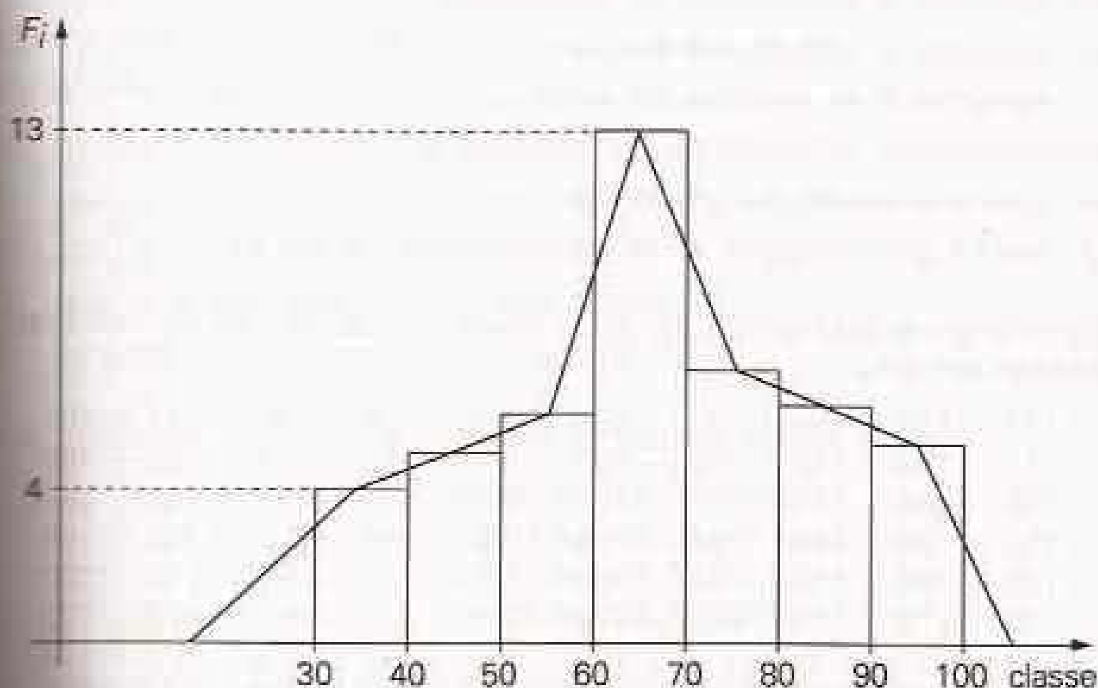
$$h \cong 64 : 7 \cong 10$$

Limite das classes, freqüências e pontos médios:

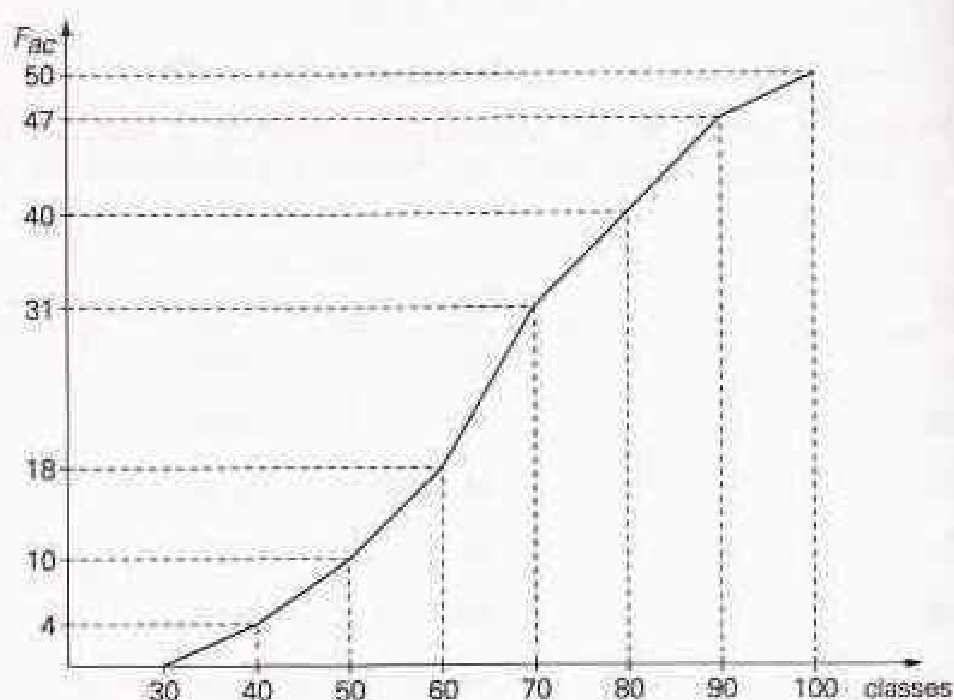
Para facilitar a contagem das freqüências, inicia-se a primeira classe por 30. Todavia, poder-se-ia iniciar com 33. Assim, a distribuição de freqüência será:

Classes	F_i	F_{ac}	f_i	X_i
30 ⊢ 40	4	4	0,08	35
40 ⊢ 50	6	10	0,12	45
50 ⊢ 60	8	18	0,16	55
60 ⊢ 70	13	31	0,26	65
70 ⊢ 80	9	40	0,18	75
80 ⊢ 90	7	47	0,14	85
90 ⊢ 100	3	50	0,06	95
Σ	50		1	

O histograma e polígono de freqüência dados por:



Assim como o gráfico das freqüências acumuladas:



EXERCÍCIOS – SÉRIE II – CAPÍTULO 5

1. Dada a amostra: 3, 4, 4, 5, 7, 6, 6, 7, 7, 4, 5, 5, 6, 6, 7, 5, 8, 5, 6, 6 pede-se:

   a) construir a distribuição de freqüência;

   b) construir o gráfico das freqüências;

   c) determinar as freqüências relativas;

   d) determinar as freqüências acumuladas;

   e) qual é a amplitude amostral;

   f) qual a porcentagem de elementos maiores que 5.

2. Considere os dados obtidos pelas medidas das alturas de 100 indivíduos (dadas em cm):

   151 – 152 – 154 – 155 – 158 – 159 – 159 – 160 – 161 – 161
   161 – 162 – 163 – 163 – 163 – 164 – 165 – 165 – 165 – 166
   166 – 166 – 166 – 167 – 167 – 167 – 167 – 167 – 168 – 168
   168 – 168 – 168 – 168 – 168 – 168 – 168 – 168 – 169 – 169
   169 – 169 – 169 – 169 – 169 – 170 – 170 – 170 – 170 – 170
   170 – 170 – 171 – 171 – 171 – 171 – 172 – 172 – 172 – 173
   173 – 173 – 174 – 174 – 174 – 175 – 175 – 175 – 175 – 176
   176 – 176 – 176 – 177 – 177 – 177 – 177 – 178 – 178 – 178
   179 – 179 – 180 – 180 – 180 – 180 – 181 – 181 – 181 – 182
   182 – 182 – 183 – 184 – 185 – 186 – 187 – 188 – 190 – 190

Pede-se determinar:

a) a amplitude amostral;
b) o número de classes;
c) a amplitude das classes;
d) os limites das classes;
e) as freqüências absolutas das classes;
f) as freqüências relativas;
g) os pontos médios das classes;
h) a freqüência acumulada;
i) o histograma – polígono de freqüência;
j) os gráficos de freqüência acumulada.

3. As notas de 32 estudantes de uma classe estão descritas a seguir:

6,0 – 0,0 – 2,0 – 6,5 – 5,0 – 3,5 – 4,0 – 7,0
8,0 – 7,0 – 8,5 – 6,0 – 4,5 – 0,0 – 6,5 – 6,0
2,0 – 5,0 – 5,5 – 5,0 – 7,0 – 1,5 – 5,0 – 5,0
4,0 – 4,5 – 4,0 – 1,0 – 5,5 – 3,5 – 2,5 – 4,5

Determinar:

a) o rol;
b) as distribuições de freqüências (variável contínua). (Sugestão: iniciar por 0 e intervalo de classe 1,5);
c) o maior e o menor graus;
d) a amplitude total;
e) qual a porcentagem dos alunos que tiveram nota menor do que 4;
f) qual o limite superior da segunda classe;
g) qual o ponto médio da quarta classe;
h) qual o ponto médio da terceira classe;
i) os gráficos (histograma e gráfico da F_{ac}).

4. Os pesos de 40 alunos estão relacionados a seguir:

69 57 72 54 93 68 72 58 64 62
65 76 60 49 74 59 66 83 70 45
60 81 71 67 63 64 53 73 81 50
67 68 53 75 65 58 80 60 63 53

a) Construir a tabela de distribuição de freqüência, dado log 40 = 1,6.
b) Construir os gráficos da distribuição.

5. Completar os dados que faltam:

Valores	F_i	F_{ac}	f_i
1	4		0,08
2	4		
3		16	0,16
4	7		0,14
5	5	28	
6		38	
7	7	45	0,14
8			

5.5 MEDIDAS DE POSIÇÕES

Foi visto nas seções anteriores a sintetização dos dados sob a forma de tabelas, gráficos e distribuições de freqüências. Aqui, destaca-se o cálculo de medidas que possibilitam representar um conjunto de dados relativos à observação de determinado fenômeno de forma resumida. São as medidas de posição. Tais medidas orientam-nos quanto à posição da distribuição no eixo x (eixo dos números reais), possibilitam comparações de séries de dados entre si pelo confronto desses números. São chamadas medidas de tendência central, pois representam os fenômenos pelos seus valores médios, em torno dos quais tendem a concentrar-se os dados.

5.5.1 Média Aritmética – Dados Não Agrupados

Sejam $x_1, x_2, ..., x_n$, portanto "n" valores da variável X. A média aritmética simples de X representada por $\bar{x}$ é definida por:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \text{ou simplesmente} \quad \bar{x} = \frac{\Sigma x}{n},$$

em que n é o número de elementos do conjunto.

Exemplo: Determinar a média aritmética simples dos valores: 3, 7, 8, 10, 11.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{3 + 7 + 8 + 10 + 11}{5}$$

$$\bar{x} = 7,8$$

5.5.2 Média Aritmética – Dados Agrupados

Quando os dados estiverem agrupados numa distribuição de freqüência usaremos a média aritmética dos valores $x_1, x_2, ..., x_n$, ponderados pelas respectivas freqüências absolutas: $F_1, F_2, F_3, ..., F_n$. Assim:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i F_i}{n} \quad \text{ou} \quad \bar{x} = \frac{\Sigma x_i F_i}{n}$$

Exemplos:

a) Dada a seguinte distribuição:

x_i	1	2	3	4
F_i	1	3	5	1

determinar a média.

Um dispositivo prático para esse cálculo é a composição da seguinte tabela:

x_i	F_i	$x_i F_i$
1	1	1
2	3	6
3	5	15
4	1	4
Σ	10	26

$$\bar{x} = \frac{\Sigma x_i F_i}{n} = \frac{26}{10} = 2{,}6$$

lembre-se de que $\Sigma F_i = n$

Exemplo: Determinar a média da distribuição:

Renda Familiar (milhares de $)	2 ⊢ 4	4 ⊢ 6	6 ⊢ 8	8 ⊢ 10	10 ⊢ 12
Nº de Famílias	5	10	14	8	3

Nesse caso, as classes são representadas pelos seus pontos médios, portanto:

Classes	F_i	x_i	$x_i F_i$
2 ├ 4	5	3	15
4 ├ 6	10	5	50
6 ├ 8	14	7	98
8 ├ 10	8	9	72
10 ├ 12	3	11	33
Σ	40		268

$$\bar{x} = \frac{\Sigma x_i F_i}{n} = \frac{268}{40}$$

$$\bar{x} = 6{,}7$$

Como a renda familiar foi dada em milhares, pode-se dizer que a renda média desse grupo de 40 famílias é de $ 6.700,00.

5.5.3 Média Geral

Sejam $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, ..., $\bar{x}_k$ as médias aritméticas de k séries e n_1, n_2, n_3, ..., n_k os números de termos destas séries, respectivamente. A média aritmética da série formada pelos termos das k séries é dada pela fórmula:

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots + n_k}$$

Exemplo: Sejam as séries:

1) 4, 5, 6, 7, 8 em que $n_1 = 5$ e $\bar{x}_1 = 6$
2) 1, 2, 3 em que $n_2 = 3$ e $\bar{x}_2 = 2$
3) 9, 10, 11, 12, 13 em que $n_3 = 5$ e $\bar{x}_3 = 11$

Então, a média geral das séries, utilizando a fórmula acima, será:

$$\bar{x} = \frac{5 \cdot 6 + 3 \cdot 2 + 5 \cdot 11}{5 + 3 + 5} = 7$$

5.5.4 Média Geométrica

Sejam x_1, x_2, x_3, ..., x_n, valores de X, associados às freqüências absolutas F_1, F_2, F_3, ..., F_n, respectivamente. A média geométrica de X é definida por:

$$Mg = \sqrt[n]{x_1^{F_1} \cdot x_2^{F_2} \cdot x_3^{F_3} \cdot \ldots \cdot x_n^{F_n}}$$

Em particular, Se $F_1 = F_2 = F_3 = \ldots = F_n = 1$, tem-se:

$$Mg = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}$$

Exemplo 1: Calcular a média geométrica dos valores 3, 6, 12, 24, 48. Logo:

$$Mg = \sqrt[5]{3 \times 6 \times 12 \times 24 \times 48}$$

$$Mg = \sqrt[5]{248.832} \therefore Mg = 12$$

Exemplo 2: Calcular a média geométrica para a distribuição:

x_i	1	2	3	5
F_i	8	6	5	3

Solução: Note que a aplicação direta da definição acarreta grande número de operações. Nestes casos é conveniente o uso dos logaritmos. Assim, tomando-se o logaritmo de ambos os membros da fórmula, tem-se:

$$\log Mg = \frac{F_1 \log x_1 + F_2 \log x_2 + F_3 \log x_3 + \ldots + F_n \log x_n}{n}$$

Logo:

$$\log Mg = \frac{8 \log 1 + 6 \log 2 + 5 \log 3 + 3 \log 5}{22}$$

$$\log Mg = \frac{8(0) + 6(0{,}3010) + 5(0{,}4771) + 3(0{,}6990)}{22}$$

$$\log Mg = 0{,}2858 \text{ ou } Mg = \text{antilog}\, 0{,}2858 = 10^{0{,}2858}$$

$$Mg = 1{,}9311$$

5.5.5 Média Harmônica

Sejam $x_1, x_2, x_3, ..., x_n$, valores de X, associados às freqüências absolutas $F_1, F_2, F_3, ..., F_n$, respectivamente.

A média harmônica de X é definida por:

$$Mh = \frac{n}{\frac{F_1}{x_1} + \frac{F_2}{x_2} + \frac{F_3}{x_3} + ... + \frac{F_n}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{F_i}{x_i}}$$

Em particular, se $F_1 = F_2 = F_3 = ... = F_n = 1$, tem-se:

$$Mh = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + ... + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Exemplo: Calcular a média harmônica para 2, 5, 8.

Então:

$$Mh = \frac{3}{\frac{1}{2} + \frac{1}{5} + \frac{1}{8}} = 3{,}64$$

EXERCÍCIOS – SÉRIE III – CAPÍTULO 5

1. Determine a média aritmética das seguintes séries:
   a) 3, 4, 1, 3, 6, 5, 6
   b) 7, 8, 8, 10, 12
   c) 3,2; 4; 0,75; 5; 2,13; 4,75
   d) 70, 75, 76, 80, 82, 83, 90

2. A média mínima para aprovação em determinada disciplina é 5,0. Se um estudante obtém as notas 7,5; 8,0; 3,5; 6,0; 2,5; 2,0; 5,5; 4,0 nos trabalhos mensais da disciplina em questão, pergunta-se se ele foi ou não aprovado.

3. Calcule para cada uma das distribuições abaixo sua respectiva média.

a)

x_i	3	4	7	8	12
F_i	2	5	8	4	3

b)

x_i	10	11	12	13
F_i	5	8	10	6

c)

x_i	F_{ac}
2	3
3	9
4	19
5	25
6	28

d)

x_i	f_i
7	1/16
8	5/18
9	1/3
10	2/9
11	5/48

e)

x_i	F_i
85	5
87	1
88	10
89	3
90	5

4. Dadas as estaturas de 140 alunos, conseguiu-se a distribuição abaixo. Calcular a média.

Estaturas (cm)	145 ⊢ 150	150 ⊢ 155	155 ⊢ 160	160 ⊢ 165	165 ⊢ 170	170 ⊢ 175	175 ⊢ 180	180 ⊢ 185
Nº dos alunos	2	10	27	38	27	21	8	7

5. Abaixo temos a distribuição dos aluguéis de 65 casas. Determine sua média.

Aluguel (milhares de $)	1,5 ⊢ 3,5	3,5 ⊢ 5,5	5,5 ⊢ 7,5	7,5 ⊢ 9,5	9,5 ⊢ 11,5
Nº de casas	12	18	20	10	5

6. Dada a distribuição

Classes	68 ⊢ 72	72 ⊢ 76	76 ⊢ 80	80 ⊢ 84
F_{ac}	8	20	35	40

determinar a média.

7. Dados os seguintes números:

1	3	5	7	9	2	4	6	8	10	15	20	25	0	[illegible]
2	3	4	5	6	7	8	9	9	8	7	8	6	5	[illegible]
3	2	1	0	10	15	20	25	12	11	8	6	4	2	[illegible]
3	5	7	9	11										

a) Construa a distribuição de freqüência do tipo "A".

b) Determine a média.

8. Turmas que possuem determinada disciplina em comum apresentam, nessa disciplina:

turma A (40 alunos) – média 6,5
turma B (35 alunos) – média 6,0
turma C (35 alunos) – média 4,0
turma D (20 alunos) – média 7,5

Determine a média geral.

9. Dada a amostra:

28	33	27	30	31	30	33	30	33	29
27	33	31	27	31	28	27	29	31	24
31	33	30	32	30	33	27	33	31	33
23	29	30	24	28	34	30	30	18	17
18	15	16	17	17	18	19	19	20	29

a) Agrupar os elementos em classes (inicie pelo 15) e use $h = 5$.

b) Construir a tabela de distribuição de freqüência do tipo "B".

c) Determinar a média.

10. Calcule a média geométrica para as séries:

a) 8, 15, 10, 12

b) 3, 4, 5, 6, 7, 8.

c)

x_i	8	9	10	11	12
F_i	12	10	7	5	3

11. Encontre a média harmônica para as séries:

a) 5, 7, 12, 15

b)

x_i	2	3	4	5	6
F_i	3	4	6	5	2

12. Tem-se $ 2.000,00 disponíveis, mensalmente, para a compra de determinado artigo que custou, nos meses de junho, julho e agosto, respectivamente, $ 200,00; $ 500,00 e $ 700,00. Qual foi o custo médio do artigo para esse período?

13. Utilizando a série de dados: 2, 7, 8 e 15, comprove as seguintes propriedades da média aritmética.

a) A soma dos desvios em torno da média é zero. Isto é, $\Sigma(x_i - \bar{x}) = 0$.

Sugestão: Calcule a média ($\bar{x}$). Depois, determine para cada valor da série o seu desvio $(x_i - \bar{x})$. Some tais desvios e a propriedade ficará provada.

b) Somando ou subtraindo a mesma quantidade arbitrária de todos os valores da série, a média ficará aumentada ou diminuída dessa mesma quantidade.

Sugestão: Utilize, por exemplo, a quantidade arbitrária 2. Calcule a média ($\bar{x}$); some o número 2 a todos os valores. Determine a média desses "novos" valores. Compare as duas médias obtidas. Observe que a segunda média supera a primeira em duas unidades, o que prova a propriedade.

c) Multiplicando ou dividindo cada termo de uma série por uma constante, a média ficará multiplicada ou dividida pela constante.

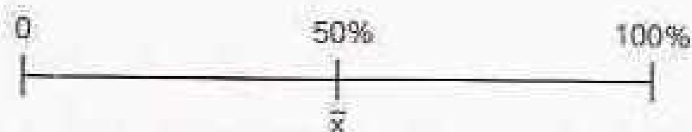
Sugestão: Admita a constante igual a 3 e prove mais essa propriedade.

d) A soma dos quadrados dos desvios medidos em relação à média é um mínimo, ou seja, é sempre menor que a soma dos quadrados dos desvios medidos em relação a outro valor qualquer. Isto é, $\Sigma(x_i - \bar{x})^2$ é mínima.

Sugestão: Utilize os desvios calculados para a propriedade a). Calcule a soma de seus quadrados e constate que ela é menor que qualquer outra soma de quadrados cuja origem não seja a média $\bar{x}$.

5.5.6 Mediana

Colocados em ordem crescente, mediana ($\tilde{x}$) é o valor que divide a amostra, ou população, em duas partes iguais. Assim:



Cálculo da mediana – variável discreta

Se n for ímpar, a mediana será o elemento central (de ordem $\frac{n+1}{2}$).
Caso n seja par, a mediana será a média entre os elementos centrais (de ordem $\frac{n}{2}$ e $\frac{n}{2} + 1$).

Exemplos: a) Dada a distribuição:

x_i	F_i	F_{ac}	
1	1	1	
2	3	4	
3	5	9	← Contém o 6º elemento
4	2	11	
Σ	11		

$n = 11$, n é ímpar, logo ($\tilde{x}$) será o elemento de ordem $\frac{n+1}{2}$, ou seja, $\frac{11+1}{2} = 6^{o}$.

Será, portanto, o sexto elemento. Para identificá-lo, abre-se a coluna da Freqüência Acumulada Crescente (F_{ac}).

Por meio dessas freqüências acumuladas encontra-se o valor (x_i) correspondente à mediana.

Neste exemplo será o 3 ($\tilde{x} = 3$). Observe: será o x_i correspondente à classe que contiver a ordem calculada.

b) Seja:

x_i	F_i	F_{ac}	
82	5	5	
85	10	15	
87	15	30	← 21º e 22º
89	8	38	
90	4	42	
Σ	42		

$n = 42$, n é par, logo ($\tilde{x}$) será a média entre os elementos de ordem $\frac{n}{2}$ e $\frac{n}{2} + 1$, ou seja, $\frac{42}{2} = 21^{o}$ e $\frac{42}{2} + 1 = 22^{o}$.

Como no exemplo anterior, identificam-se os elementos de ordem 21 e 22 pela F_{ac}.

Assim: 21º corresponde a 87.

22º corresponde a 87, logo,

$\tilde{x} = \frac{87 + 87}{2} = 87$.

Cálculo da mediana – variável contínua.

1º Passo: Calcula-se a ordem $\frac{n}{2}$. Como a variável é contínua, não se preocupe se n é par ou ímpar.

2º Passo: Pela F_{ac} identifica-se a classe que contém a mediana (classe Md).

3º Passo: Utiliza-se a fórmula:

$$\tilde{x} = \ell_{Md} + \frac{\left(\frac{n}{2} - \Sigma f\right) \cdot h}{F_{Md}}$$

em que

ℓ_{Md} = limite inferior da classe Md

n = tamanho da amostra ou número de elementos

Σf = soma das freqüências anteriores à classe Md

h = amplitude da classe Md

F_{Md} = freqüência da classe Md.

Exemplo: Dada a distribuição amostral, calcular a mediana.

Classes	F_i	F_{ac}	
35 ⊢ 45	5	5	
45 ⊢ 55	12	17	
55 ⊢ 65	18	35	← classe Md
65 ⊢ 75	14	49	
75 ⊢ 85	6	55	
85 ⊢ 95	3	58	
Σ	58		

1º Passo: Calcula-se $\frac{n}{2}$. Como $n = 58$, temos $\frac{58}{2} = 29^{\circ}$.

2º Passo: Identifica-se a classe Md pela F_{ac}. Neste caso, a classe Md é a 3ª.

3º Passo: Aplica-se a fórmula:

$$\tilde{x} = \ell_{Md} + \frac{\left(\frac{n}{2} - \Sigma f\right) \cdot h}{F_{Md}}$$

No caso

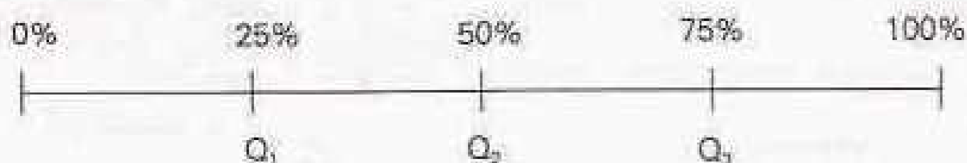
$$\ell_{Md} = 55; \quad n = 58; \quad \Sigma f = 17; \quad h = 10; \quad F_{Md} = 18.$$

Logo:

$$\tilde{x} = 55 + \frac{\left(\frac{58}{2} - 17\right) \cdot 10}{18} = 61{,}67$$

5.5.7 Quartis

Os quartis dividem um conjunto de dados em quatro partes iguais. Assim:



Q_1 = 1º quartil, deixa 25% dos elementos.

Q_2 = 2º quartil, coincide com a mediana, deixa 50% dos elementos.

Q_3 = 3º quartil, deixa 75% dos elementos.

Eis as fórmulas para os cálculos de Q_1 e Q_3 para o caso de variáveis contínuas.

Determinação de Q_1:

1º Passo: Calcula-se $\frac{n}{4}$.

2º Passo: Identifica-se a classe Q_1 pela F_{ac}.

3º Passo: Aplica-se a fórmula:

$$Q_1 = \ell_{Q_1} + \frac{\left(\frac{n}{4} - \Sigma f\right) \cdot h}{F_{Q_1}}$$

Determinação de Q_3:

1º Passo: Calcula-se $\frac{3n}{4}$

2º Passo: Identifica-se a classe Q_3 pela F_{ac}.

3º Passo: Aplica-se a fórmula:

$$Q_3 = \ell_{Q_3} + \frac{\left(\frac{3n}{4} - \Sigma f\right) \cdot h}{F_{Q_3}}$$

Exemplo: Dada a distribuição, determinar os quartis (Q_1 e Q_3) e mediana.

Classes	F_i	F_{ac}	
7 ⊢ 17	6	6	
17 ⊢ 27	15	21	→ classe Q_1
27 ⊢ 37	20	41	→ classe *Md*
37 ⊢ 47	10	51	→ classe Q_3
47 ⊢ 57	5	56	
Σ	56		

1º Passo: $n = 56$

Q_1 ? $\tilde{x}$? Q_3 ?

$\frac{n}{4} = \frac{56}{4} = 14^{\text{º}}$ $\frac{n}{2} = \frac{56}{2} = 28^{\text{º}}$ $\frac{3n}{4} = 42^{\text{º}}$

2º Passo: Pela F_{ac} identifica-se a classe Q_1, classe *Md* e classe Q_3

3º Passo: Uso das fórmulas

Para Q_1 temos: $\ell_{Q_1} = 17$, $n = 56$, $\Sigma f = 6$; $h = 10$, $F_{Q_1} =$

Para $\tilde{x}$ temos: $\ell_{MD} = 27$, $n = 56$, $\Sigma f = 21$, $h = 10$, $F_{Md} =$

Para Q_3 temos: $\ell_{Q_3} = 37$, $n = 56$, $\Sigma f = 41$, $h = 10$, $F_{Q_3} =$

Logo:

$$Q_1 = 17 + \frac{\left(\frac{56}{4} - 6\right) \cdot 10}{15} = 22{,}33$$

$$\tilde{x} = 27 + \frac{\left(\frac{56}{2} - 21\right) \cdot 10}{20} = 30{,}5$$

$$Q_3 = 37 + \frac{\left(\frac{3 \cdot 56}{4} - 41\right) \cdot 10}{10} = 38$$

Diante desses resultados, pode-se afirmar que, nesta distribuição, tem-se



Isto é:

22,33 deixa 25% dos elementos

30,5 deixa 50% dos elementos

38 deixa 75% dos elementos

5.5.8 Decis

Continuando o estudo das medidas separatrizes mediana e quartis, tem-se os decis. São os valores que dividem a série em 10 partes iguais.



Como você já deve ter percebido, a fórmula neste caso também é semelhante às separatrizes anteriores. Ei-la:

1º Passo: Calcula-se $\frac{i \cdot n}{10}$, em que i = 1, 2, 3, 4, 5, 6, 7, 8 e 9

2º Passo: Identifica-se a classe D_i pela F_{ac}

3º Passo: Aplica-se a fórmula:

$$D_i = \ell_{D_i} + \frac{\left(\frac{in}{10} - \Sigma f\right) \cdot h}{F_{D_i}}$$

em que ℓ_{D_i} = limite inferior da classe D_i, i = 1, 2, 3, ..., 9

n = tamanho da amostra

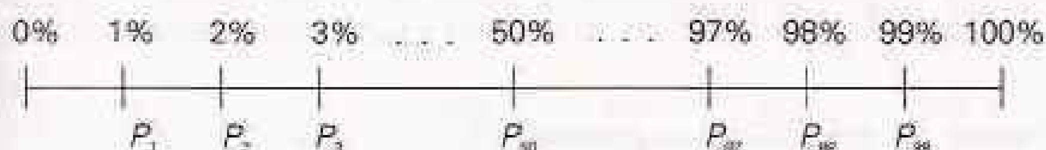
h = amplitude da classe D_i

F_{D_i} = freqüência da classe D_i

Σ_f = soma das freqüências anteriores à classe D_i

5.5.9 Percentis

São as medidas que dividem a amostra em 100 partes iguais. Assim:



O cálculo de um percentil é dado por:

1º Passo: Calcula-se $\frac{in}{100}$, em que $i = 1, 2, 3, ..., 98, 99$.

2º Passo: Pela F_{ac} identifica-se a classe P_i.

3º Passo: Usa-se a fórmula:

$$P_i = \ell_{P_i} + \frac{\left(\frac{in}{100} - \Sigma f\right) \cdot h}{F_{P_i}}$$

em que ℓ_{P_i} = limite da classe P_i, em que $i = 1, 2, 3, ..., 99$

n = tamanho da amostra

Σf = soma das freqüências anteriores à classe P_i

h = amplitude da classe P_i

F_{P_i} = freqüência da classe P_i

Exemplo: Determinar o 4º Decil e o 72º Percentil da seguinte distribuição:

	Classes	F_i	F_{ac}	
	4 ⊢ 9	8	8	
	9 ⊢ 14	12	20	← classe D_4
	14 ⊢ 24	17	37	← classe P_{72}
	19 ⊢ 24	3	40	
Σ		40		

Cálculo do D_4:

1º Passo:

$$\frac{in}{10} = \frac{4 \cdot 40}{10} = 16^{\circ}$$

Cálculo do P_{72}:

1º Passo:

$$\frac{in}{100} = \frac{72 \cdot (40)}{100} = 28{,}8^{\circ}$$

2º Passo: Identifica-se a classe D_4 e P_{72} pela F_{ac}.

3º Passo:

Para D_4: $\ell_{D_4} = 9$; $\Sigma f = 8$; $n = 40$; $h = 5$; $F_{D_4} = 12$

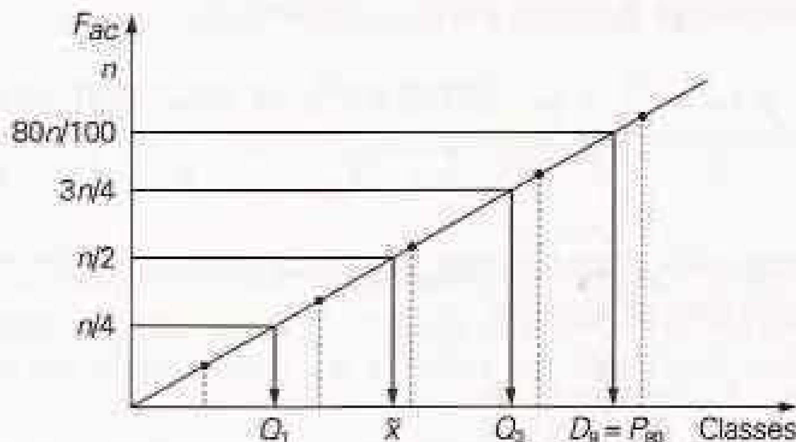
Para P_{72}: $\ell_{P_{72}} = 14$; $\Sigma f = 20$; $n = 40$; $h = 5$; $F_{P_{72}} = 17$

$$D_4 = 9 + \frac{\left(\frac{4 \cdot 40}{10} - 8\right) \cdot 5}{12} = 12{,}33$$

$$P_{72} = 14 + \frac{\left(\frac{72 \cdot 40}{100} - 20\right) \cdot 5}{17} = 16{,}89$$

Portanto, nesta distribuição, o valor 12,33 divide a amostra em duas partes: uma com 40% dos elementos e a outra com 60% dos elementos. O valor 16,89 indica que 72% da distribuição estão abaixo dele e 28% acima.

Para a determinação das separatrizes (mediana, quartis, decis e percentis) pode-se utilizar o gráfico da freqüência acumulada. Assim:



5.5.10 Moda

Dentre as principais medidas de posição, destaca-se a Moda. É o valor mais freqüente da distribuição. Para distribuições simples (sem agrupamento em classes), a identificação da Moda é facilitada pela simples observação do elemento que apresenta maior freqüência. Assim, para a distribuição

x_i	243	245	248	251	307
F_i	7	17	23	20	8

a Moda será 248. Indica-se Mo = 248.

Para dados agrupados em classes, há diversas fórmulas para o cálculo da Moda.

a) 1º processo: fórmula de Czuber

1º Passo:

Identifica-se a classe modal (aquela que possuir maior freqüência).

2º Passo:

Aplica-se a fórmula:

$$Mo = \ell + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h$$

em que:

- ℓ = limite inferior da classe modal
- Δ_1 = diferença entre a freqüência da classe modal e a imediatamente anterior
- Δ_2 = diferença entre a freqüência da classe modal e a imediatamente posterior
- h = amplitude da classe.

Exemplo: Determinar a moda para a distribuição.

Classes	0 ⊢ 1	1 ⊢ 2	2 ⊢ 3	3 ⊢ 4	4 ⊢ 5	Σ
F_i	3	10	17	8	5	43

1º Passo: Indica-se a classe modal. No caso, trata-se da 3ª classe 2 ⊢ 3.

2º Passo: Aplica-se a fórmula:

$$Mo = \ell + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h$$

em que:

$\ell = 2$

$\Delta_1 = 17 - 10 = 7$

$\Delta_2 = 17 - 8 = 9$

$h = 1$

Portanto:

$$Mo = 2 + \frac{7}{7 + 9} \cdot 1 = 2{,}44$$

Exemplo: Calcular a moda para a distribuição.

Salários (US$)	80 ⊢ 180	180 ⊢ 250	250 ⊢ 300	300 ⊢ 500
Nº de empregados	70	140	140	60

Observe que as amplitudes das classes não são iguais. Nestes casos é preciso calcular as "densidades" das classes: $F_i \div h$ para se identificar qual a classe modal (aquela com maior densidade). Assim:

Salários (US$)	F_i	F_i/h	
80 $\vdash^{(100)}$ 180	70	70/100 = 0,7	
180 $\vdash^{(70)}$ 250	140	140/70 = 2,0	
250 $\vdash^{(50)}$ 300	140	140/50 = 2,8	→ classe modal
300 $\vdash^{(200)}$ 500	60	60/200 = 0,3	

1º Passo: Classe modal. No caso será a 3ª classe 250 ⊢ 300.

2º Passo: Aplica-se a fórmula, onde:

$\ell = 250$
$\Delta_1 = 2{,}8 - 2{,}0 = 0{,}8$
$\Delta_2 = 2{,}8 - 0{,}3 = 2{,}5$
$h = 50$

$$Mo = 250 + \frac{0{,}8}{0{,}8 + 2{,}5} \cdot 50 = 262{,}12$$

Portanto, o salário mais freqüente é US$ 262,12.

2º processo: fórmula de Pearson

$$Mo = 3\tilde{x} - 2\bar{x}$$

Ou seja, a moda é aproximadamente a diferença entre o triplo da mediana e o dobro da média. Esta fórmula dá uma boa aproximação quando a distribuição apresenta razoável simetria em relação à média.

EXERCÍCIOS – SÉRIE IV – CAPÍTULO 5

1. Para cada série, determine a mediana:

I) 1, 3, 3, 4, 5, 6, 6
II) 1, 3, 3, 4, 6, 8, 8, 9
III) 12, 7, 10, 8, 8
IV) 82, 86, 88, 84, 91, 93

2. Para cada distribuição, determine a mediana:

I)

x_i	2	3	4	5	7
F_i	3	5	8	4	2

II)

x_i	73	75	77	79	81
F_i	2	10	12	5	2

III)

x_i	12	13	15	17
F_{ac}	5	13	18	20

IV)

x_i	232	235	237	240
F_{ac}	15	40	55	61

3. Para cada distribuição, determine a mediana:

I)

Classes	**$1 \vdash 3$**	**$3 \vdash 5$**	**$5 \vdash 7$**	**$7 \vdash 9$**	**$9 \vdash 11$**	**$11 \vdash 13$**
F_i	3	5	8	6	4	3

II)

Classes	**$22 \vdash 25$**	**$25 \vdash 28$**	**$28 \vdash 31$**	**$31 \vdash 34$**
F_i	18	25	30	20

4. Para cada série, determine a moda:

I) 3, 4, 7, 7, 7, 8, 9, 10

II) 43, 40, 42, 43, 47, 45, 45, 43, 44, 48

5. Para cada distribuição, determine a moda:

I)

x_i	72	75	78	80
F_i	8	18	28	38

II)

x_i	2,5	3,5	4,5	6,5
F_i	7	17	10	5

6. Para cada distribuição, determine a moda pelos dois processos:

I)

Classes	**$7 \vdash 10$**	**$10 \vdash 13$**	**$13 \vdash 16$**	**$16 \vdash 19$**	**$19 \vdash 22$**
F_i	6	10	15	10	5

II)

Classes	**$10 \vdash 20$**	**$20 \vdash 30$**	**$30 \vdash 40$**	**$40 \vdash 50$**
F_{ac}	7	19	28	32

7. Para as distribuições:

I)

Classes	**$4 \vdash 6$**	**$6 \vdash 8$**	**$8 \vdash 10$**	**$10 \vdash 12$**
F_i	4	11	15	5

calcule D_6, P_{65} e Q_1.

II)

Classes	20 ⊢ 30	30 ⊢ 40	40 ⊢ 50	50 ⊢ 60	60 ⊢ 70
F_{ac}	3	8	18	22	24

calcule D_2, P_{43} e Q_3.

8. Abaixo temos a distribuição do número de acidentes por dia, durante 53 dias, em certa rodovia:

Nº de acidentes	0	1	2	3	4
Nº de dias	20	15	10	5	3

Pede-se:

a) determinar a média;

b) determinar a mediana;

c) calcular a moda;

d) qual a porcentagem de dias em que tivemos dois ou mais acidentes por dia?

9. O número de operários acidentados por mês, numa fábrica, nos últimos dois anos, foi:

Ano \ Mês	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
1975	4	8	3	6	7	7	3	8	2	4	3	3
1976	7	4	6	5	10	5	4	3	5	4	4	1

Faça $X \rightarrow$ número de operários acidentados por mês.

a) construa a distribuição de freqüência (tipo A).

b) Calcule a média, mediana e moda.

10. Sendo:

Idade (anos)	10 ⊢ 14	14 ⊢ 18	18 ⊢ 22	22 ⊢ 26	26 ⊢ 30	30 ⊢ 34	34 ⊢ 38	38 ⊢ 42
Nº de pessoas	15	28	40	30	20	15	10	5

a) determinar a média

b) calcular a medida que deixa 50% dos elementos;

c) determinar a moda (fórmula de Czuber);

d) calcular o 3º decil;

e) determinar a medida que deixa 1/4 dos elementos;

f) calcular o percentil 80;

g) qual a porcentagem das pessoas maiores de idade?

11. Foi pedido aos alunos de uma classe de 40 alunos que escolhessem um dentre os números 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9. Obteve-se o seguinte resultado:

8 – 0 – 2 – 3 – 3 – 5 – 7 – 7 – 7 – 9
8 – 4 – 1 – 9 – 6 – 6 – 6 – 8 – 3 – 3
7 – 7 – 6 – 0 – 1 – 3 – 3 – 3 – 7 – 7
6 – 5 – 5 – 1 – 2 – 5 – 2 – 5 – 3 – 2

a) montar a distribuição de freqüência (tipo A);

b) determinar a média;

c) qual foi o nº mais escolhido? O que ele representa?

d) calcule a mediana.

12. Abaixo estão dadas as notas (em créditos) de 50 alunos:

60 85 33 52 65 77 84 65 74 57
71 35 81 50 35 64 74 47 54 68
80 61 41 91 55 73 59 53 77 45
41 55 78 48 69 85 67 39 60 76
94 98 66 66 73 42 65 94 88 89

Pede-se:

a) determinar a amplitude total da amostra;

b) número de classes pela fórmula Sturges. Dado log 50 = 1,7;

c) amplitude das classes;

d) quais as classes? (inicie pelo 30);

e) freqüências absolutas das classes;

f) freqüências relativas;

g) pontos médios das classes;

h) freqüência acumulada;

i) histograma;

j) polígono de freqüência;

k) gráfico da freqüência acumulada;

l) média

m) moda – processo gráfico;

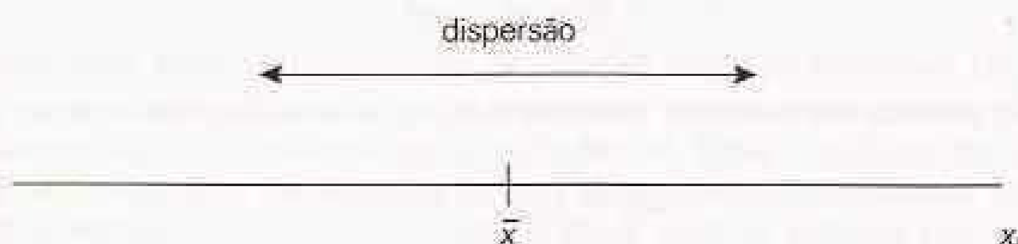
n) mediana – pelo gráfico do item k;

o) 1º e 3 º quartis – pelo gráfico do item k;

p) 7º decil e 55º percentil pelo gráfico.

5.6 MEDIDAS DE DISPERSÃO

São medidas estatísticas utilizadas para avaliar o grau de variabilidade, ou dispersão, dos valores em torno da média. Servem para medir a representatividade da média.



Sejam as séries:

a) 20, 20, 20

b) 15, 10, 20, 25, 30

Tem-se: $\bar{x}_a = 20$ e $\bar{x}_b = 20$

Observe: Apesar de as séries terem médias iguais, na série "a" não se tem dispersão, enquanto os valores da série "b" apresentam dispersões em torno da média 20. Assim, a média é muito mais representativa para a série "a" do que para a série "b".

5.6.1 Amplitude Total

É a diferença entre o maior e o menor dos valores da série.

$$R = x_{máx} - x_{mín}$$

Exemplo: Para a série 10, 12, 20, 22, 25, 33, 38

$$R = 38 - 10 = 28$$

A utilização da amplitude total como medida de dispersão é muito limitada, pois, sendo uma medida que depende apenas dos valores externos, é instável, não sendo afetada pela dispersão dos valores internos.

5.6.2 Desvio Médio

Desde que se deseja medir a dispersão os dados em relação à média, parece interessante a análise dos desvios em torno da média. Isto é, analisar os desvios:

$$d_i = (x_i - \overline{x})$$

Mas a soma de todos os desvios é igual a zero. Isto é:

$$\Sigma d_i = \Sigma(x_i - \overline{x}) = 0$$

Logo, será preciso encontrar uma maneira de se trabalhar com os desvios sem que a soma dê zero. Dessa forma define-se desvio médio como:

$$D_M = \frac{\Sigma |d_i| F_i}{n} = \frac{\Sigma |x_i - \overline{x}| F_i}{n}$$

Veja que os desvios foram considerados em módulo, evitando-se assim que soma fosse nula. Observe exemplo adiante.

5.6.3 Variância

Neste caso considera-se o quadrado de cada desvio $(x_i - \overline{x})^2$, evitando com isso que $\Sigma d_i = 0$. Assim, a definição da variância populacional é dada por:

$$\sigma^2 = \frac{\Sigma(x_i - \overline{X})^2 \cdot F_i}{N} = \frac{\Sigma d_i^2 \cdot F_i}{N}$$

Trata-se da média aritmética dos quadrados dos desvios.

Observação: 1. σ^2 indica variância populacional e lê-se sigma ao quadrado.

2. $\overline{X}$ da fórmula é a média da população.

Para o caso do cálculo da variância amostral é conveniente o uso da seguinte fórmula:

$$S^2 = \frac{\Sigma(x_i - \overline{x})^2 \cdot F_i}{n - 1}$$

Como você deve ter notado, as diferenças entre as fórmulas são: para o caso da variância populacional: σ^2, utiliza-se a média populacional ($\overline{X}$) tendo como denominador o tamanho da população: N. Para o cálculo da variância amostral: S^2, utiliza-se a média amostral ($\overline{x}$), tendo como denominador o tamanho da amostra menos um: $(n - 1)$. Fórmulas práticas para os cálculos das variâncias:

$$\sigma^2 = \frac{1}{N}\left[\Sigma X_i^2 F_i - \frac{(\Sigma X_i F_i)^2}{N}\right]$$

ou

$$S^2 = \frac{1}{n - 1}\left[\Sigma x_i^2 F_i - \frac{(\Sigma x_i F_i)^2}{n}\right]$$

que foram obtidas por transformações nas respectivas fórmulas originais.

5.6.4 Desvio-padrão

Observando-se a fórmula original para o cálculo da variância, nota-se que é uma soma de quadrados. Dessa forma, se a unidade da variável for, por exemplo, metro (m) teremos como resultado metro ao quadrado (m^2). Para se ter a unidade original, necessita-se definir outra medida de dispersão, que é a raiz quadrada da variância – o desvio-padrão. Assim:

$$\sigma = \sqrt{\sigma^2}$$ é o desvio-padrão populacional.

$$S = \sqrt{S^2}$$ é o desvio-padrão amostral

Resumindo: para o cálculo do desvio-padrão deve-se primeiramente determinar o valor da variância e, em seguida, extrair a raiz quadrada desse resultado.

Exemplo: Calcular o desvio-médio, a variância e o desvio-padrão da segu[illegible] distribuição amostral:

x_i	5	7	8	9	11
F_i	2	3	5	4	2

1º) *Cálculo do desvio médio:*

$$D_M = \frac{\Sigma\,|x_i - \bar{x}| \cdot F_i}{n} \quad \text{ou} \quad D_M = \frac{\Sigma\,|d_i| \cdot F_i}{n}$$

Primeiramente precisa-se do valor da média:

	x_i	F_i	$x_i F_i$
	5	2	10
	7	3	21
	8	5	40
	9	4	36
	11	2	22
Σ		16	129

$$\bar{x} = \frac{\Sigma x_i F_i}{n} = \frac{129}{16} = 8{,}06$$

Para o cálculo do D_M são abertas novas colunas, assim:

x_i	F_i	$x_i F_i$	$\lvert x_i - \bar{x} \rvert = \lvert d_i \rvert$	$\lvert d_i \rvert F_i$
5	2	10	$\lvert 5 - 8{,}06 \rvert = 3{,}06$	6,12
7	3	21	$\lvert 7 - 8{,}06 \rvert = 1{,}06$	3,18
8	5	40	$\lvert 8 - 8{,}06 \rvert = 0{,}06$	0,30
9	4	36	$\lvert 9 - 8{,}06 \rvert = 0{,}94$	3,76
11	2	22	$\lvert 11 - 8{,}06 \rvert = 2{,}94$	5,88
Σ	16	129		19,24

Portanto,

$$D_M = \frac{\Sigma\,|d_i| \cdot F_i}{n} = \frac{19{,}24}{16} = 1{,}20$$

2º) *Cálculo da variância amostral*

$$S^2 = \frac{1}{n-1}\left[\Sigma x_i^2 F_i - \frac{(\Sigma x_i F_i)^2}{n}\right]$$

Observe que o cálculo será facilitado, pois sabe-se que: $n = 16$; $\Sigma x_i F_i = 129$. Resta encontrar $\Sigma x_i^2 F_i$. Para tanto, uma nova coluna é considerada na tabela.

	x_i	F_i	$x_i F_i$	$x_i^2 F_i$
	5	2	10	50
	7	3	21	147
	8	5	40	320
	9	4	36	324
	11	2	22	242
Σ		16	129	1.083

Observe:

$x_1^2 F_1 = 5^2 \cdot 2 = 50$

$x_2^2 F_2 = 7^2 \cdot 3 = 147$

$\vdots \quad \vdots \quad \vdots \quad \vdots$

$$S^2 = \frac{1}{n-1}\left[\Sigma x_i^2 F_i - \frac{(\Sigma x_i F_i)^2}{n}\right] = \frac{1}{16-1}\left[1.083 - \frac{(129)^2}{16}\right]$$

$$S^2 = \frac{1}{15}\left[1.083 - \frac{16.641}{16}\right] = \frac{1}{15}\left[\frac{17.328 - 16.641}{16}\right]$$

$$S^2 = \frac{687}{15 \cdot 16} = 2{,}86$$

Logo, a variância amostral é 2,86; ou seja, $S^2 = 2{,}86$.

3º) *Cálculo do desvio-padrão amostral*

Como $S = \sqrt{S^2}$, logo $S = \sqrt{2{,}86} = 1{,}69$.

Resumindo: A distribuição possui média 8,06. Isto é, seus valores estão em torno de 8,06 e seu grau de concentração é de 1,2, medido pelo Desvio Médio, e de 1,69, medido pelo Desvio-padrão.

Eis outro exemplo:

Dada a distribuição amostral abaixo, encontrar a média, o desvio médio e o desvio-padrão.

Classes	2 ├ 4	4 ├ 6	6 ├ 8	8 ├ 10	10 ├ 12
F_i	2	4	7	4	3

A constituição da tabela auxiliar para os cálculos deve ser construída à medida que você for necessitando dos resultados parciais, portanto, não se preocupe em memorizar a ordem das colunas. Eis a tabela auxiliar:

	Classes	F_i	x_i	$x_i F_i$	$\lvert x_i - \bar{x} \rvert$	$\lvert d_i \rvert F_i$	$x_i^2 F_i$
	2 ├ 4	2	3	6	$\lvert 3 - 7{,}2 \rvert = 4{,}2$	8,4	18
	4 ├ 6	4	5	20	$\lvert 5 - 7{,}2 \rvert = 2{,}2$	8,8	100
	6 ├ 8	7	7	49	$\lvert 7 - 7{,}2 \rvert = 0{,}2$	1,4	343
	8 ├ 10	4	9	36	$\lvert 9 - 7{,}2 \rvert = 1{,}8$	7,2	324
	10 ├ 12	3	11	33	$\lvert 11 - 7{,}2 \rvert = 3{,}8$	11,4	363
Σ		20		144		37,2	1.148

$$\bar{x} = \frac{144}{20} = 7{,}2 \qquad D_M = \frac{37{,}2}{20} = 1{,}86$$

$$S^2 = \frac{1}{20 - 1}\left[1.148 - \frac{(144)^2}{20}\right] = 5{,}86$$

$$S = \sqrt{5{,}85} = 2{,}42$$

Exemplo: Cálculo da variância populacional. Determinar a variância para a série

x_i	2	3	5	6	7
F_i	1	4	5	3	2

Solução: A fórmula prática para o cálculo da variância populacional é dada por:

$$\sigma^2 = \frac{1}{N}\left[\Sigma x_i^2 F_i - \frac{(\Sigma x_i F_i)^2}{N}\right]$$

Será conveniente a construção da seguinte tabela:

	x_i	F_i	$x_i F_i$	$x_i^2 F_i$
	2	1	2	4
	3	4	12	36
	5	5	25	125
	6	3	18	108
	7	2	14	98
Σ		15	71	371

Assim

$$\sigma^2 = \frac{1}{15}\left[371 - \frac{(71)^2}{15}\right]$$

$$\sigma^2 = 2{,}33$$

Quanto ao desvio-padrão populacional, tem-se:

$$\sigma = \sqrt{2{,}33}$$

$$\sigma = 1{,}53$$

5.6.5 Coeficiente de Variação

Trata-se de uma medida relativa de dispersão útil para a comparação em termos relativos do grau de concentração em torno da média de séries distintas. É dado por:

$$CV = \frac{\sigma}{\bar{X}} \times 100 \quad \text{ou} \quad CV = \frac{S}{\bar{x}} \times 100$$

onde: σ = desvio-padrão populacional

$\bar{X}$ = média populacional

S = desvio-padrão amostral

$\bar{x}$ = média amostral.

O coeficiente de variação é expresso em porcentagens.

Exemplo: Numa empresa, o salário médio dos homens é de \$ 4.000,00, com desvio-padrão de \$ 1.500,00, e o das mulheres é em média de \$ 3.000,00, com desvio-padrão de \$ 1.200,00. Então:

$$\text{para os homens } C \cdot V = \frac{\sigma}{\bar{x}} = \frac{1.500}{4.000} \times 100 = 37{,}5\%$$

$$\text{para as mulheres } C \cdot V = \frac{\sigma}{\bar{x}} = \frac{1.200}{3.000} \times 100 = 40\%$$

Logo, podemos concluir que os salários das mulheres apresentam maior dispersão relativa que os dos homens.

Diz-se que a distribuição possui pequena variabilidade (dispersão) quando o coeficiente der até 10%; média dispersão quando estiver acima de 10% até 20%; e grande dispersão quando superar 20%. Alguns analistas consideram:

baixa dispersão: $CV \leq 15\%$

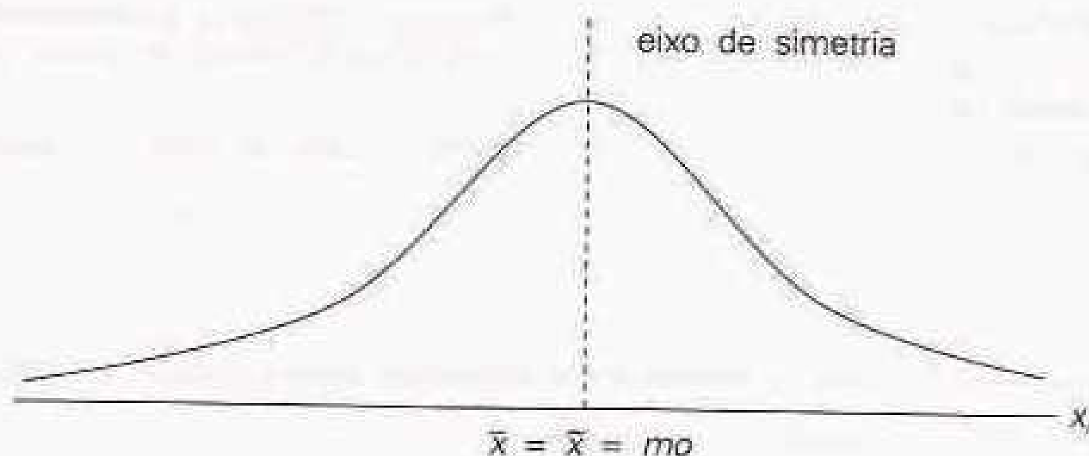
média dispersão: $15\% < CV < 30\%$

alta dispersão: $CV \geq 30\%$

5.7 MEDIDAS DE ASSIMETRIA

Denomina-se assimetria o grau de afastamento de uma distribuição da unidade de simetria.

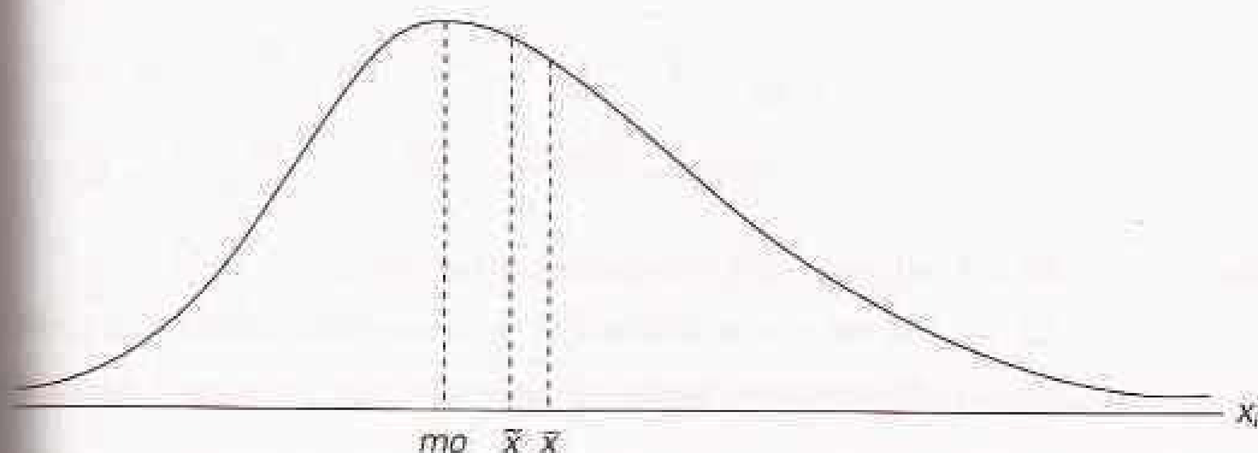
Em uma distribuição simétrica tem-se igualdade dos valores da média, mediana e moda. Eis um exemplo gráfico de distribuição simétrica.



Em uma distribuição assimétrica positiva, ou assimétrica à direita, tem-se:

$Mo < \tilde{x} < \bar{x}$

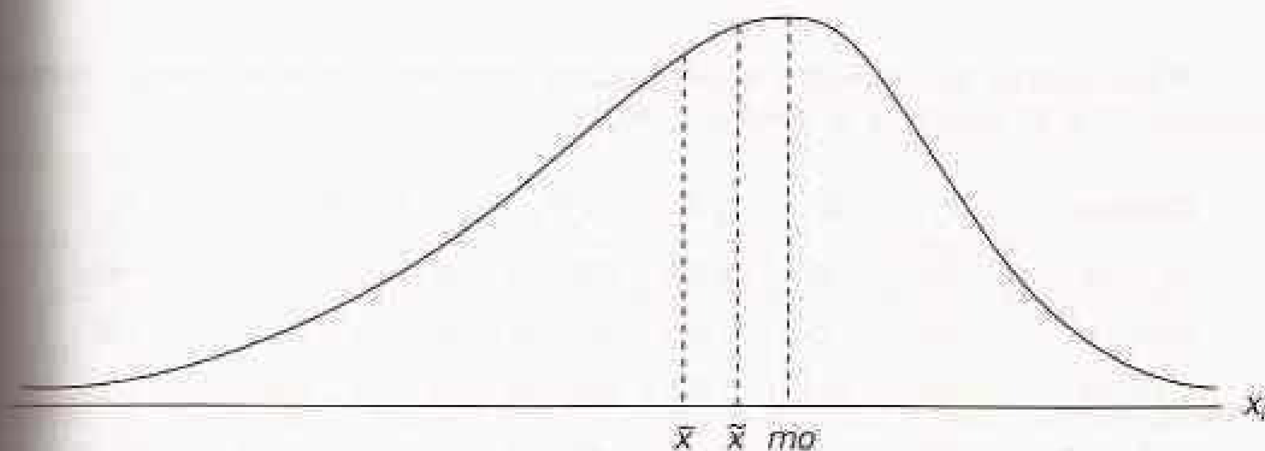
Eis um exemplo gráfico de distribuição assimétrica positiva:



Em uma distribuição assimétrica negativa, ou assimétrica à esquerda, tem-se:

$$\bar{x} < \tilde{x} < Mo$$

Eis um exemplo de distribuição assimétrica negativa:



Existem várias fórmulas para o cálculo do coeficiente de assimetria, dentre elas são úteis:

1º *Coeficiente de Pearson*

$$AS = \frac{\bar{x} - Mo}{\sigma} \quad \text{ou} \quad AS = \frac{\bar{x} - Mo}{S}$$

2º *Coeficiente de Pearson*

$$AS = \frac{Q_1 + Q_3 - 2\tilde{x}}{Q_3 - Q_1}$$

Se: AS = 0 diz-se que a distribuição é simétrica

AS > 0 diz-se que a distribuição é assimétrica positiva (à dir

AS < 0 diz-se que a distribuição é assimétrica negativa (à esqu

Pode-se utilizar qualquer uma das fórmulas para identificar o gra assimetria de uma distribuição.

Exemplo: Dada a distribuição amostral. Calcular os dois coeficientes de Pea

Salários ($ 1.000,00)	**30 ⊢ 50**	**50 ⊢ 100**	**100 ⊢ 150**
Empregados	80	50	30

Para aplicar as fórmulas é necessário calcular a média, moda, des padrão, 1º e 3º quartis e a mediana. Assim:

Classes	F_i	x_i	$x_i F_i$	$x_i^2 F_i$	$F_i \div h$	F_{ac}
30 ⊢ 50	80	40	3.200	128.000	80 ÷ 20 = 4	80
50 ⊢ 100	50	75	3.750	281.250	50 ÷ 50 = 1	130
100 ⊢ 150	30	125	3.750	468.750	30 ÷ 50 = 0,6	160
Σ	160		10.700	878.000		

$$\bar{x} = \frac{10.700}{160} = 66,875$$

$$Mo = 30 + \frac{4}{4 + 3} \cdot 20 = 41,429$$

$$S^2 = \frac{1}{159}\left[878.000 - \frac{(10.700)^2}{160}\right] = 1021,62$$

$$S = 31,96$$

$$Q_1 = 30 + \frac{(40 - 0) \cdot 20}{80} = 40$$

$$Q_3 = 50 + \frac{(120 - 80) \cdot 50}{50} = 90$$

$$\tilde{x} = 30 + \frac{(80 - 0) \cdot 20}{80} = 50$$

$$AS = \frac{\bar{x} - Mo}{S} = \frac{66,875 - 41,429}{31,96} = 0,796$$

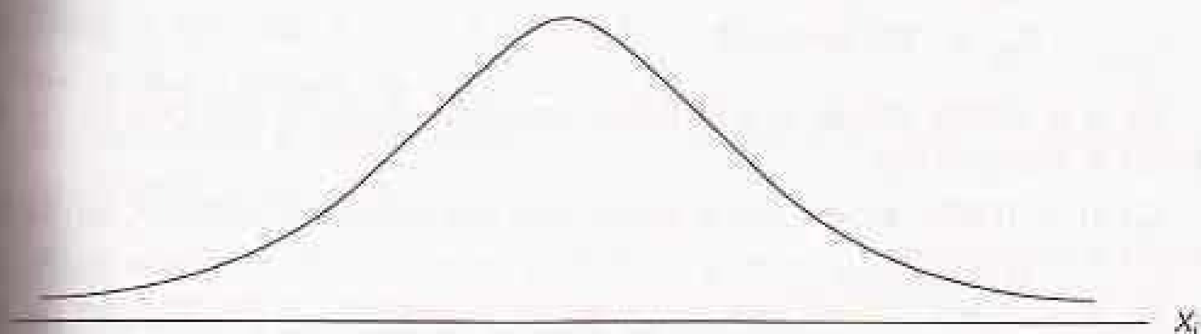
$$AS = \frac{Q_1 + Q_3 - 2\tilde{x}}{Q_3 - Q_1} = \frac{40 + 90 - 2 \cdot (50)}{90 - 40} = 0,6$$

Como, nos dois casos, AS > 0, diz-se que a distribuição é assimétrica positiva.

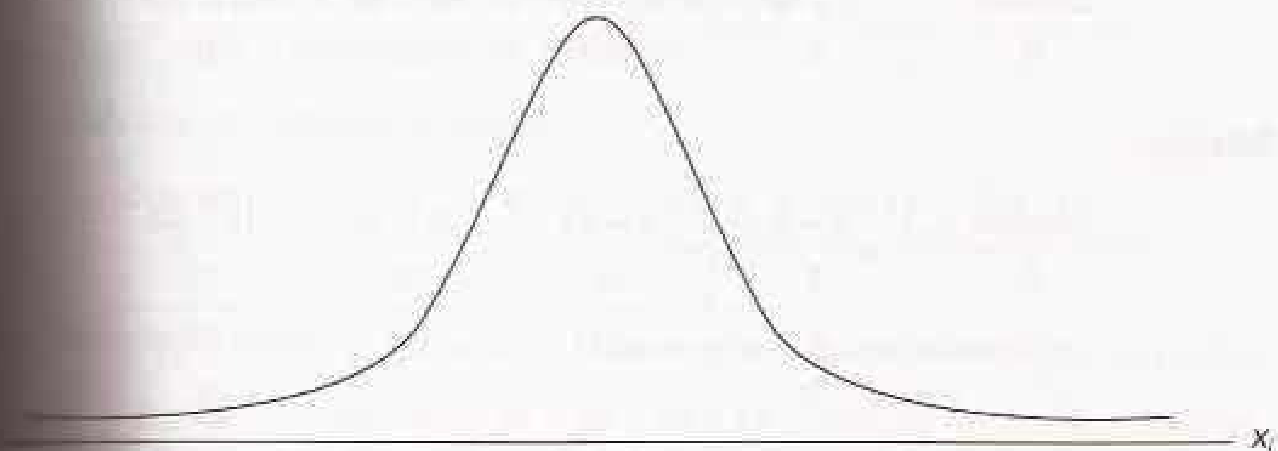
5.8 MEDIDAS DE CURTOSE

Denomina-se curtose o grau de achatamento da distribuição.

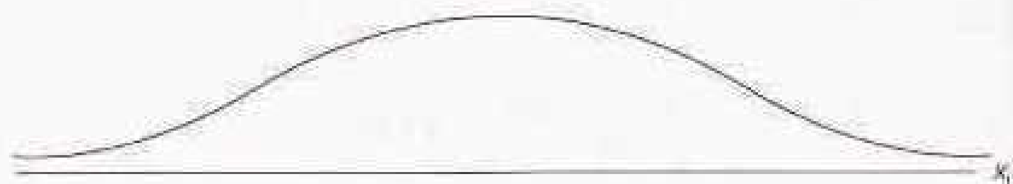
Uma distribuição nem chata, nem delgada, chama-se mesocúrtica. Eis um exemplo gráfico de distribuição mesocúrtica.



Uma distribuição delgada chama-se leptocúrtica. Eis um exemplo gráfico de distribuição leptocúrtica.



Uma distribuição achatada denomina-se platicúrtica. Eis um exemplo gráfico de distribuição platicúrtica.



Para medir o grau de curtose utiliza-se o coeficiente:

$$K = \frac{Q_3 - Q_1}{2\,(P_{90} - P_{10})}$$

em que: Q_3 = 3º quartil
Q_1 = 1º quartil
P_{90} = 90º percentil
P_{10} = 10º percentil

Se $K = 0{,}263$, diz-se que a curva correspondente à distribuição de freqüência é mesocúrtica.

Se $K > 0{,}263$, diz-se que a curva correspondente à distribuição de freqüência é platicúrtica.

Se $K < 0{,}263$, diz-se que a curva correspondente à distribuição de freqüência é leptocúrtica.

Exemplo: Dizer que tipo de curva corresponde à distribuição amostral.

Classes	**3 ⊢ 8**	**8 ⊢ 13**	**13 ⊢ 18**	**18 ⊢ 23**
Fi	5	15	20	10

Solução:

Classes	**3 ⊢ 8**	**8 ⊢ 13**	**13 ⊢ 18**	**18 ⊢ 23**
Fi	5	15	20	10
F_{ac}	5 ↑ classe P_{10}	20 ↑ classe Q_1	40 ↑ classe Q_3	50 ↑ classe P_{90}

$$Q_1 = 8 + \frac{(12,5 - 5) \cdot 5}{15} = 10,5$$

$$Q_3 = 13 + \frac{(37,5 - 20) \cdot 5}{20} = 17,38$$

$$P_{10} = 3 + \frac{(5 - 0) \cdot 5}{5} = 8$$

$$P_{90} = 18 + \frac{(45 - 40) \cdot 5}{10} = 20,5$$

$$K = \frac{17,38 - 10,5}{2(20,5 - 8)} = 0,27.$$

Portanto, $K > 0,263$, logo a curva correspondente é suavemente platicúrtica.

EXERCÍCIOS – SÉRIE V – CAPÍTULO 5

Medidas de Dispersão, Assimetria, Curtose.

1. Dada a amostra: 2, 3, 4, 5, 7, 10, 12
   a) qual é a amplitude total?
   b) determine o desvio médio;
   c) calcule a variância.

2. Para a série 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 9, 9.
   a) construir a distribuição simples de freqüência;
   b) calcular a amplitude;
   c) determinar o desvio médio;
   d) calcular a variância populacional;
   e) determinar o desvio-padrão populacional;
   f) calcular o coeficiente de variação.

3. Calcular a variância amostral:

Classes	$2 \vdash 4$	$4 \vdash 6$	$6 \vdash 8$	$8 \vdash 10$	$10 \vdash 12$
F_i	3	5	8	6	3

4. Num teste aplicado a 20 alunos, obteve-se a seguinte distribuição de pontos:

Pontos	$35 \vdash 45$	$45 \vdash 55$	$55 \vdash 65$	$65 \vdash 75$	$75 \vdash 85$	$85 \vdash 95$
Nº de alunos	1	3	8	3	3	2

a) calcular o desvio médio;
b) determinar a variância populacional;
c) determinar o desvio-padrão;
d) calcular o coeficiente de variação;
e) determinar o coeficiente de assimetria (1º coeficiente de Pearson);
f) calcular o coeficiente de curtose.

5. Abaixo temos a distribuição de freqüência dos pesos de uma amostra de 45 alunos:

Peso em k	40 ⊢ 45	45 ⊢ 50	50 ⊢ 55	55 ⊢ 60	60 ⊢ 65	65 ⊢ 70
Nº de alunos	4	10	15	8	5	3

a) determinar a média;
b) determinar a variância;
c) qual é o valor do coeficiente de variação?
d) a distribuição é simétrica?
e) a distribuição é mesocúrtica?

6. Sendo:

Classes	30 ⊢ 40	40 ⊢ 50	50 ⊢ 60	60 ⊢ 70	70 ⊢ 80
F_i	10	20	35	25	10

calcular $\overline{x}$, S^2, S, CV, AS e K.

7. A distribuição abaixo possui desvio-padrão igual a 3,02. Determine o valor do coeficiente de variabilidade.

Classes	0 ⊢ 4	4 ⊢ 8	8 ⊢ 12
Freqüência	2	3	2

8. Um fabricante de caixas de cartolina fabrica três tipos de caixa. Testa-se a resistência de cada caixa, tomando-se uma amostra de 100 caixas e determinando-se a pressão necessária para romper cada caixa. São os seguintes os resultados dos testes:

Tipos de caixas	*A*	*B*	*C*
Pressão média de ruptura (bária)	150	200	300
Desvio-padrão das pressões (bária)	40	50	60

a) que tipo de caixa apresenta a menor variação absoluta na pressão de ruptura?

b) que tipo de caixa apresenta a maior variação relativa na pressão de ruptura?

9. Um pesquisador da rádio XY aborda 30 transeuntes ao acaso e pergunta-lhes a idade. O resultado é dado pela tabela:

35	26	39	25	39	22
42	40	39	22	21	40
16	32	39	21	28	39
18	37	23	14	27	44
30	32	21	15	26	43

a) resuma as informações sob forma de uma distribuição de freqüência. Dado log 30 = 1,48;

b) apresente os dados na forma de um histograma;

c) calcule a média e o desvio-padrão amostral.

10. É dada a distribuição dos salários semanais de 100 funcionários:

Salário por semana ($)	500 ├ 1.000	1.000 ├ 1.500	1.500 ├ 2.000	2.000 ├ 2.500	2.500 ├ 3.000
Nº de empregados	26	43	17	9	5

a) calcule a variância populacional;

b) a distribuição é assimétrica?

c) a distribuição é leptocúrtica?

EXERCÍCIOS – SÉRIE VI – CAPÍTULO 5

Medidas de Posição, Dispersão, Assimetria e Curtose

1. Dada a série: 1,2; 1,4; 1,5; 1,8; 2 calcular a média e o desvio-padrão populacional.

2. Calcular:

a) média

b) mediana

c) moda

d) desvio médio

e) coeficiente de assimetria da seguinte distribuição:

Altura (cm)	*Freqüência*
$160 \vdash 164$	5
$164 \vdash 168$	13
$168 \vdash 172$	22
$172 \vdash 176$	25
$176 \vdash 180$	10
$180 \vdash 184$	3

3. Num fim de semana, o supermercado X vendeu as seguintes quantidades de carne:

Tipo de Carne	*Preço ($ por kg)*	*Quantidade (kg)*
Boi	35	1.000
Porco	38	450
Galinha	39	600
Peru	45	350
Peixe	28	250

Qual foi o preço médio por quilograma vendido?

4. Completar os dados que faltam para a seguinte distribuição:

x_i	F_i	F_{ac}	f_i
1	4		0,04
2	8		
3		30	0,18
4	27		0,27
5	15	72	
6		83	
7	10	93	0,10
8			

5. Encontrar a freqüência correspondente à terceira classe da distribuição a seguir, sabendo-se que a média é igual a 11,50.

X_i	5	8	13	18	25
F_i	4	5	....	3	1

6. Achar o 1º quartil, o 7º decil e o 73º percentil da distribuição:

Classes	0 ⊢ 1	1 ⊢ 2	2 ⊢ 3	3 ⊢ 4	4 ⊢ 5
F_i	10	12	12	10	6

7. Obter a moda e a variância para a distribuição amostral:

Classes	0 ⊢ 25	25 ⊢ 50	50 ⊢ 75	75 ⊢ 100	100 ⊢ 125
F_i	20	140	180	40	10

8. Lançando um dado 50 vezes, obteve-se a seguinte distribuição:

x_i	F_i
1	6
2	11
3	6
4	7
5	9
6	11

calcular a variância populacional e o desvio-padrão.

9. Calcule a média e a variância amostral:

x_i	30.000	30.002	30.004	30.006	30.008	30.010
F_{ac}	10	22	36	46	50	52

10. Estudar a distribuição abaixo, com respeito à assimetria e à curtose.

Classes	150 ⊢ 200	200 ⊢ 250	250 ⊢ 300	300 ⊢ 350	350 ⊢ 400	400 ⊢ 450	450 ⊢ 500
F_i	5	16	21	28	19	8	3

11. Cronometrando o tempo para várias provas de uma gincana automobilística, encontramos:

Equipe 1: 40 provas
tempo médio: 45 segundos
variância: 400 segundos ao quadrado

Equipe 2:	tempo:	20	40	50	80
	nº de provas:	10	15	30	5

a) qual o coeficiente de variação relativo à equipe 1?
b) qual a média da equipe 2?
c) qual o desvio-padrão relativo à equipe 2?
d) qual a média aritmética referente às duas equipes consideradas em conjunto?
e) qual a equipe que apresentou resultados mais homogêneos? Justifique.

12. Dada a amostra de 60 rendas (em milhares) de dada região geográfica:

10	7	8	5	4	3	2	9	9	6
3	15	1	13	14	4	3	6	6	8
10	11	12	13	14	2	15	5	4	10
2	1	3	8	10	11	13	14	15	16
8	9	5	3	2	3	3	4	4	4
5	6	7	8	9	1	12	13	14	16

a) agrupar os elementos em classes. Sendo $K = 6$ e $h = 3$;
b) construir o histograma e o polígono de freqüência acumulada;
c) construir o gráfico de freqüência acumulada;
d) calcular a média;
e) calcular a mediana;
f) determinar o 3º quartil;
g) calcular o 4º decil;
h) calcular o 47º percentil;
i) determinar a medida que deixa 25% das rendas;
j) calcular o desvio médio;
l) determinar a variância;
m) determinar o desvio-padrão;
n) qual é o valor do coeficiente de variação?
o) a distribuição é simétrica?

p) a distribuição é mesocúrtica?

q) usando o gráfico da freqüência acumulada, determine o 1º quartil, o 7º decil e o 80º percentil;

r) prepare um relatório para a descrição das rendas dessas famílias.

EXERCÍCIOS – SÉRIE VII – CAPÍTULO 5

Para cada uma das questões abaixo, assinale a alternativa correta.

1. A média aritmética é a razão entre:

   a) ( ) o número de valores e o somatório deles;
   b) ( ) o somatório dos valores e o número deles;
   c) ( ) os valores extremos;
   d) ( ) os dois valores centrais.

2. Na série 60, 90, 80, 60, 50 a moda será:

   a) ( ) 50; b) ( ) 60;
   c) ( ) 66; d) ( ) 90.

3. A medida que tem o mesmo número de valores abaixo e acima dela é:

   a) ( ) a moda; b) ( ) a média;
   c) ( ) a mediana; d) ( ) o lugar mediano.

4. A soma dos desvios entre cada valor e a média é:

   a) ( ) positiva; b) ( ) negativa;
   c) ( ) diferente de zero; d) ( ) zero.

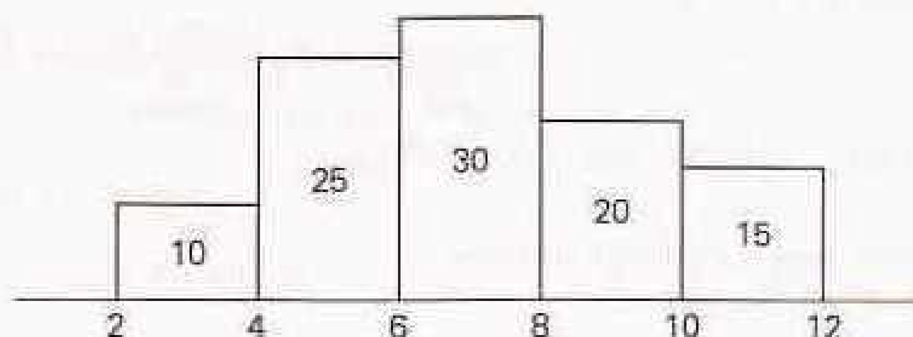
5. Na série 60, 50, 70, 80, 90 o valor 70 será:

   a) ( ) a média e a moda;
   b) ( ) a média e a mediana;
   c) ( ) a mediana e a moda;
   d) ( ) a média, a mediana e a moda.

6. Quando queremos verificar a questão de uma prova que apresentou maior número de erros, utilizamos:

   a) ( ) moda; b) ( ) média;
   c) ( ) mediana; d) ( ) qualquer das anteriores.

7. Dado o histograma abaixo, no interior de cujos retângulos foram anotadas as freqüências absolutas, então a mediana é:

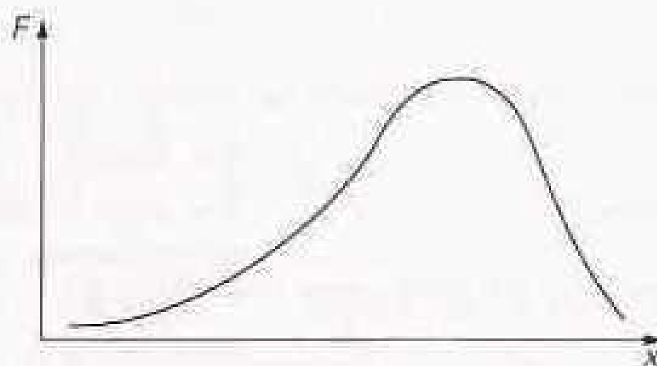


a) ( ) 6,5; b) ( ) 8,0; c) ( ) 7,5; d) ( ) 7,0.

8. Na série 15, 20, 30, 40, 50, há abaixo da mediana:

a) ( ) 3 valores; b) ( ) 2 valores;
c) ( ) 3,5 valores; d) ( ) 4 valores.

9. Dada a figura a seguir, podemos afirmar que:



a) ( ) a moda é maior do que a mediana e menor do que a média;
b) ( ) a moda é menor do que a mediana e maior do que a média;
c) ( ) a moda é menor do que a mediana e esta maior do que a média;
d) ( ) a mediana é maior do que a média e menor do que a moda.

10. O coeficiente de variação é uma medida que expressa a razão entre:

a) ( ) desvio-padrão e média;
b) ( ) média e desvio-padrão;
c) ( ) amplitude semi-interquartélica e mediana;
d) ( ) desvio-padrão e moda.

11. O cálculo da variância supõe o conhecimento da:

a) ( ) média;
b) ( ) mediana;
c) ( ) ponto médio;
d) ( ) moda.

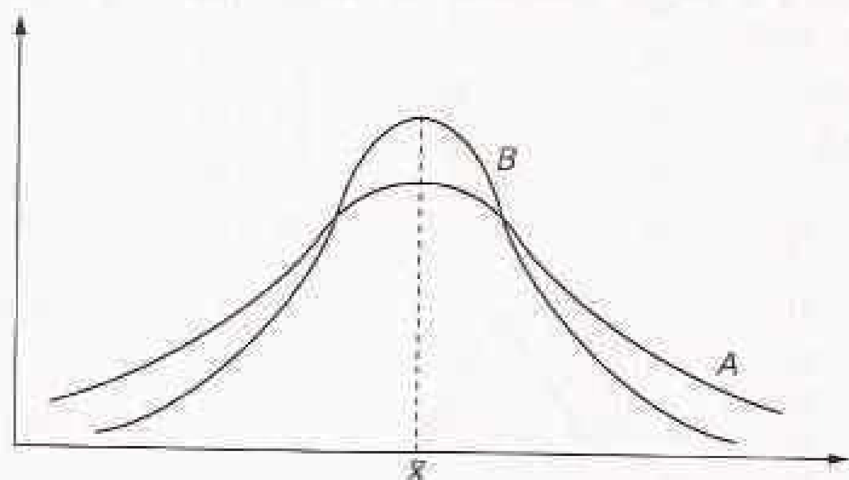
12. Numa distribuição de valores iguais, o desvio-padrão é:

a) ( ) negativo;
b) ( ) positivo;
c) ( ) a unidade;
d) ( ) zero.

13. Na série 10, 20, 40, 50, 70, 80, a mediana será:

a) ( ) 30;
b) ( ) 35;
c) ( ) 40;
d) ( ) 45.

14. Examinando a figura a seguir podemos dizer:



a) ( ) o desvio-padrão da distribuição *A* é maior do que o da distribuição *B*, e as médias são iguais:
b) ( ) o desvio-padrão de *A* é menor do que o de *B* e as médias são diferentes;
c) ( ) o desvio-padrão de *A* é igual ao de *B*, independentemente do valor da média;
d) ( ) as distribuições possuem o mesmo coeficiente de variação.

15. Realizou-se uma prova de matemática para duas turmas. Os resultados foram os seguintes:

Turma *A*: $\bar{x} = 5$ e $\sigma = 2,5$

Turma *B*: $\bar{x} = 4$ e $\sigma = 2$

Com esses resultados, podemos afirmar:

a) ( ) a turma *B* apresentou maior dispersão absoluta;
b) ( ) a dispersão relativa é igual à dispersão absoluta;

c) ( ) tanto a dispersão absoluta quanto a relativa são maiores para a turma *B*;

d) ( ) a dispersão absoluta de *A* é maior do que a de *B*, mas em termos relativos as duas turmas não diferem quanto ao grau de dispersão das notas.

16. O desvio-padrão de um conjunto de dados é 9. A variância será:

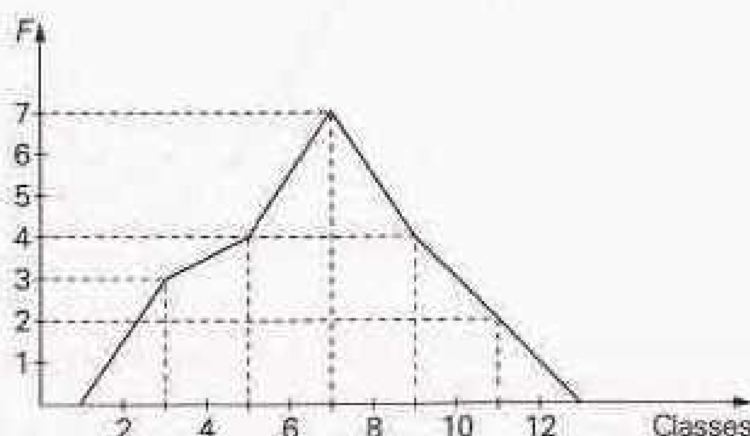
a) ( ) 3; b) ( ) 18;
c) ( ) 36; d) ( ) 81.

17. 50% dos dados da distribuição situam-se:

a) ( ) abaixo da média; b) ( ) acima da mediana;
c) ( ) abaixo da moda; d) ( ) acima da média.

18. Dada a figura a seguir (polígono de freqüência), o primeiro quartil da distribuição será:

a) ( ) 5,0; b) ( ) 5,5;
c) ( ) 4,8; d) ( ) 3,0.



19. Os coeficientes de variação dos resultados abaixo são:

Estatística: $\bar{x} = 80$; $S = 16$

História: $\bar{x} = 20$; $S = 5$

a) ( ) 16% e 40% b) ( ) 20% e 25%;
c) ( ) 50% e 40%; d) ( ) 80% e 40%.

20. Média, mediana e moda são medidas de:

a) ( ) dispersão; b) ( ) posição;
c) ( ) assimetria; d) ( ) curtose.

21. Uma empresa possui dois serventes recebendo salários de $ 2.500,00 cada um, quatro escriturários recebendo $ 6.000,00 cada um, um chefe de escritório com salário de $ 10.000,00 e três técnicos recebendo $ 22.000,00 cada um. A média destes salários é:

a) ( ) $ 1.050,00 b) ( ) $ 5.050,00;
c) ( ) $ 26.250,00 d) ( ) n.r.a.

22. O valor dominante de uma distribuição de freqüência chama-se:

a) ( ) mediana; b) ( ) média;
c) ( ) moda; d) ( ) 1º quartil.

23. Na distribuição abaixo:

Classes	*Freqüência*
30 ⊢ 40	10
40 ⊢ 50	20
50 ⊢ 60	35
60 ⊢ 70	25
70 ⊢ 80	10

A moda é:

a) ( ) 50,6; b) ( ) 55;
c) ( ) 50; d) ( ) 56.

24. Para a distribuição:

Classes	150 ⊢ 200	200 ⊢ 250	250 ⊢ 300	300 ⊢ 350	350 ⊢ 400	400 ⊢ 450	450 ⊢ 500
F_i	5	16	21	28	19	8	3

A média será:

a) ( ) 350; b) ( ) 313;
c) ( ) 324,76; d) ( ) 323,80.

25. O valor da medida que deixa 45% dos elementos da distribuição:

Renda	10 ⊢ 20	20 ⊢ 30	30 ⊢ 40	40 ⊢ 50	50 ⊢ 60	60 ⊢ 70	70 ⊢ 80	80 ⊢ 90	90 ⊢ 100
Nº de famílias	50	100	150	250	150	100	80	70	50

é:

a) ( ) 45; b) ( ) 50;
c) ( ) 46; d) ( ) 63.

26. O 5º decil da distribuição:

Classes	2 ⊢ 4	4 ⊢ 6	6 ⊢ 8	8 ⊢ 10	10 ⊢ 12
F_i	5	7	10	3	5

é:

a) ( ) 7,20; b) ( ) 5,50;
c) ( ) 6,60; d) ( ) 7,20.

27. A média da distribuição:

Classes	0 ⊢ 6	6 ⊢ 12	12 ⊢ 18
F_{ac}	1	2	5

é:

a) ( ) 12,0; b) ( ) 8,5;
c) ( ) 10,83 d) ( ) 11,4.

28. O desvio médio da distribuição:

Classes	90 ⊢ 110	110 ⊢ 130	130 ⊢ 150
F_i	2	1	2

é:

a) ( ) 12; b) ( ) 14;
c) ( ) 16; d) ( ) 18.

29. A variância da distribuição:

Classes	1 ⊢ 3	3 ⊢ 5	5 ⊢ 7
f_i	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$

é:

a) ( ) 2,24; b) ( ) 2,8;
c) ( ) 2,5; d) ( ) 4.

30. A média de uma série de valores iguais a uma constante é:

a) ( ) zero;
b) ( ) o valor da constante;
c) ( ) a unidade;
d) ( ) não é possível calcular o desvio-padrão.

EXERCÍCIOS – SÉRIE VIII – CAPÍTULO 5

1. Explique qual a utilidade das medidas de posição. Dê 3 exemplos.

2. O que são medidas de dispersão?

3. Fale sobre as medidas de curtose.

4. Se multiplicarmos todos os elementos de uma série por uma constante, que acontecerá com a média? E com a variância da série?

5. Quanto vale $\Sigma (x_i - \overline{x})$?

6. Se somarmos a todos os elementos de uma série um número, o que acontecerá com a média e a variância da série?

7. O 1º decil é igual ao décimo percentil? Explique.

8. Para analisar os dados de uma folha de pagamentos, quais medidas você utilizaria para:

a) descobrir o salário mais freqüente;
b) descobrir o salário que divide os pagamentos em partes iguais;
c) descobrir a dispersão absoluta em torno da média;
d) descobrir o grau de dispersão relativo.

9. Numa distribuição, teremos sempre a mediana e a média entre o 1º e 3º quartis. Discuta.

6 DISTRIBUIÇÕES AMOSTRAIS

6.1 INTRODUÇÃO

No Capítulo 4 foram apresentados os principais modelos de distribuição de probabilidade, enquanto no Capítulo 5 foram destacadas as medidas que caracterizam uma amostra.

Neste capítulo juntam-se os modelos e as medidas obtendo-se as distribuições amostrais dos principais estimadores.

O conhecimento dessas distribuições é a base para aplicar as técnicas de inferências estatísticas apresentadas nos capítulos seguintes.

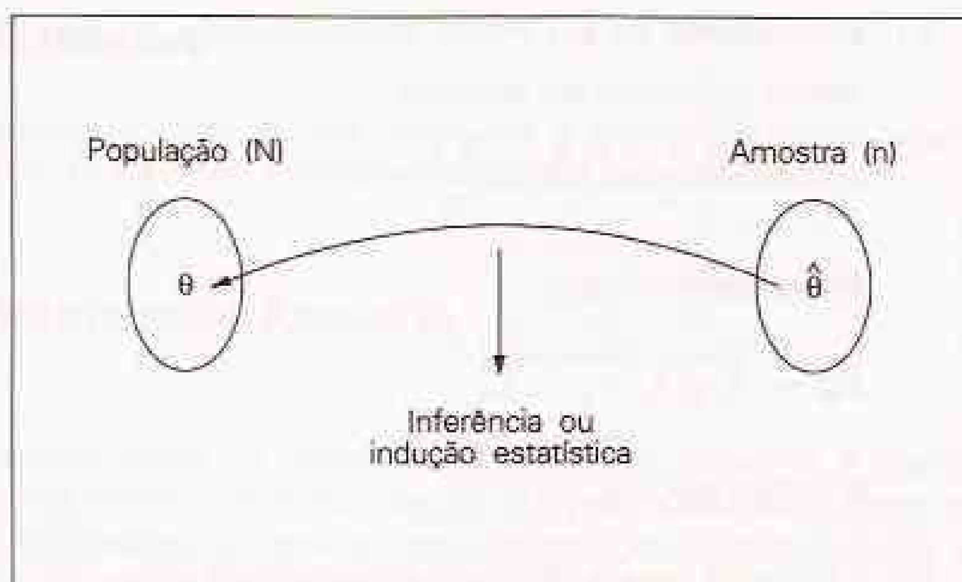
6.2 PRINCIPAIS CONCEITOS

6.2.1 Inferência ou Indução Estatística

Trata-se do processo de obter informações sobre uma população a partir de resultados observados na amostra.

De modo geral tem-se uma população com grande número de elementos, e deseja-se, a partir de uma amostra dessa população, conhecer "o mais próximo possível" algumas características da população.

Seja X uma das variáveis da população que se deseja estudar. Seja θ uma característica (medida) de X que se quer conhecer. Esse "conhecimento" de θ se dá pela construção de um estimador $\hat{\theta}$ que revelará o valor mais aproximado de θ a partir dos elementos amostrais. Assim:



Para efeitos do desenvolvimento dos demais conceitos serão consideradas populações infinitas. Ou seja, amostragens de populações muito grandes podem ser consideradas como amostragens de populações infinitas. Portanto, os parâmetros populacionais θ são desconhecidos.

6.2.2 Amostra Aleatória

Seja X uma variável populacional que se deseja estudar. Uma amostra aleatória de X é o conjunto de n variáveis aleatórias independentes $(X_1, X_2, \ldots X_n)$ tal que cada X_i ($i = 1, 2, \ldots, n$) tem a mesma característica, ou distribuição, da variável X.

Assim, se $X \overset{d}{=} N(\mu; \sigma^2)$, cada X_i terá distribuição $N(\mu; \sigma^2)$.

6.2.3 Estimador ou Estatística

Dada uma amostra aleatória, estimador ou estatística é qualquer variável aleatória função dos elementos amostrais. Ou seja:

$\hat{\theta} = f(x_1, x_2, \ldots, x_n)$ é um estimador de θ

Observe que a definição é extremamente elástica, permitindo que "qualquer" combinação das variáveis amostrais $(X_1, X_2, \ldots, X_n)$ seja um estimador ou estatística.

As medidas de posição, dispersão, assimetria e curtose vistas no Capítulo 5 são exemplos de estimadores. Assim:

a) Estimadores para a média populacional: μ.

Média aritmética ou amostral

$$\bar{x} = \frac{x_1, x_2 + \ldots + x_n}{n} = \frac{\Sigma x_i}{n}$$

Média geométrica

$$Mg = \sqrt[n]{x_1, x_2, x_3, \ldots x_n}$$

e outras médias.

b) Estimador para a variância populacional: σ^2

Variância amostral

$$S^2 = \frac{\Sigma (x_i - \bar{x})^2 F_i}{n - 1}$$

c) Estimador para a proporção ou probabilidade de um evento populacional P.

Freqüência relativa

$$f = \hat{p} = \frac{x}{n} = \frac{\text{número de casos favoráveis ao evento}}{\text{número total de casos}}$$

d) Estimador para o desvio-padrão populacional σ.

Desvio-padrão amostral

$$S = \sqrt{S^2}$$

e) Estimador para a soma, ou diferença, de duas médias populacionais $(\mu_1 \pm \mu_2)$.

Estimadores: a) $(\bar{x}_1 + \bar{x}_2)$

b) $Mg_1 \pm Mg_2$

e outras médias.

Pelo que foi apresentado deve-se ter um critério para escolher o "melhor" estimador de um parâmetro populacional. Numa das próximas seções desse capítulo você conhecerá alguns critérios para se "eleger" um bom estimador.

6.2.4 Estimativa

O valor numérico de um estimador é conhecido como uma estimativa. Assim, $\bar{x} = 17{,}8$ é uma estimativa da média populacional μ.

6.2.5 Distribuição Amostral

Considere todas as possíveis amostras de tamanho n que podem ser extraídas de determinada população. Se para cada uma delas se calcular um valor do estimador, tem-se uma distribuição amostral desse estimador. Ora, como o estimador é uma variável aleatória, pode-se determinar suas características, isto é, encontrar sua média, variância, desvio-padrão...

As distribuições amostrais são fundamentais para o processo de inferência estatística, como poderá ser constatado nos capítulos posteriores.

6.3 DISTRIBUIÇÃO AMOSTRAL DAS MÉDIAS

Lembrando o conceito de distribuição amostral, visto anteriormente, busca-se descobrir qual é a distribuição da média aritmética $\bar{x}$?

Sabe-se que $\bar{x} = \dfrac{\Sigma x_i}{n}$ (média aritmética) é um estimador da média populacional μ. O estimador $\bar{x}$ é uma variável aleatória, portanto, busca-se conhecer sua distribuição de probabilidade.

6.3.1 Teorema 1

A média da distribuição amostral das médias, denotada por $\mu(\bar{x})$ é igual à média populacional μ. Isto é:

$$E[\bar{x}] = \mu(\bar{x}) = \mu$$

Assim, é provado que a média das médias amostrais é igual à media populacional.

6.3.2 Teorema 2

Se a população é infinita, ou se a amostragem é com reposição, então a variância da distribuição amostral das médias, denotada por $\sigma^2(\bar{x})$ é dada por:

$$E[(\bar{x} - \mu)^2] = \sigma^2(\bar{x}) = \frac{\sigma^2}{n}$$

onde σ^2 é a variância da população. Isto é, pode-se afirmar que, para populações infinitas, ou amostragens com reposição, a variância da distribuição das médias é igual à variância da população dividida pelo tamanho da amostra.

6.3.3 Teorema 3

Se a população é finita, ou se a amostragem é sem reposição, então a variância da distribuição amostral das médias é dada por:

$$\sigma^2(\bar{x}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

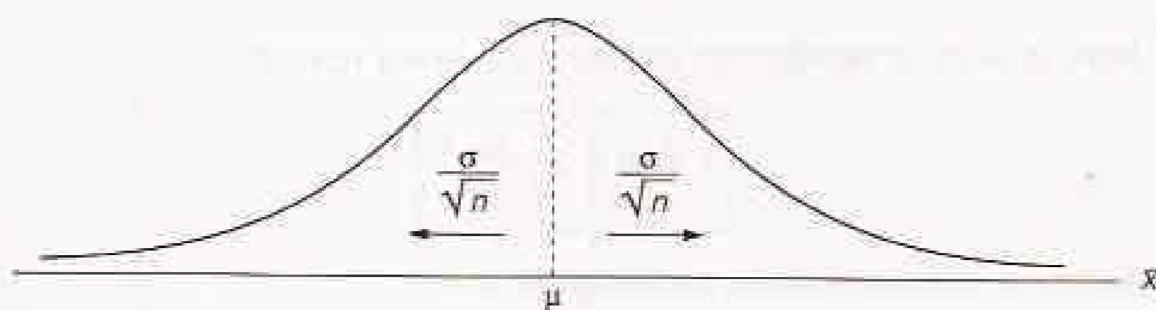
sendo que: $\mu(\bar{x}) = \mu$

6.3.4 Teorema 4

Se a população tem ou não distribuição normal com média μ e variância σ^2, então a distribuição das médias amostrais será normalmente distribuída com média μ e variância $\frac{\sigma^2}{n}$.

Esses quatro teoremas provam que a média amostral $(\bar{x})$ tem distribuição normal com média igual a média da população (μ) e variância dada por $\frac{\sigma^2}{n}$ para populações infinitas, e $\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ para populações finitas.

Graficamente:



Ou ainda:

$$\bar{x} \overset{d}{=} N\left(\mu; \frac{\sigma^2}{n}\right) \text{ ou } \bar{x} \overset{d}{=} N\left(\mu; \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)\right)$$

Com distribuições padronizadas dadas por:

$$Z_i = \frac{\bar{x}_i - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{ou} \quad Z_i = \frac{\bar{x}_i - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}}$$

6.4 DISTRIBUIÇÃO AMOSTRAL DAS FREQÜÊNCIAS RELATIVAS

Seja X uma população infinita, e p a probabilidade (ou proporção) e certo evento de X. Logo $1 - p = q$ será a probabilidade de o evento não ocorrer.

Seja $(x_1, x_2, \ldots x_n)$ uma amostra aleatória de n elementos dessa população e x o número de sucessos na amostra. É fácil identificar como uma variável aleatória com distribuição Binomial (número de sucessos na amostra), de média np e variância npq.

Então, a distribuição amostral da freqüência relativa $\hat{p} = f = \frac{x}{n}$ será dada por:

$$E[f] = E\left[\frac{x}{n}\right] = \frac{np}{n} = p$$

$$\text{Var}[f] = \text{Var}\left[\frac{x}{n}\right] = \frac{npq}{n^2} = \frac{pq}{n}$$

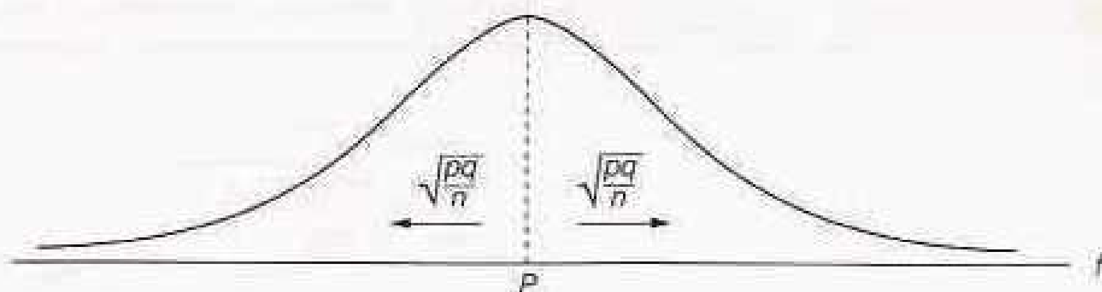
Para $n \geqslant 30$ a distribuição amostral de f será normal:

$$\boxed{f \overset{d}{=} N\left(p; \frac{pq}{n}\right)}$$

Assim, a sua distribuição padronizada será:

$$Z_i = \frac{f_i - p}{\sqrt{\frac{pq}{n}}}$$

Ou, graficamente:



6.5 DISTRIBUIÇÃO AMOSTRAL DE VARIÂNCIAS

Sabe-se que a variância da população é designada por σ^2.

Seja S^2 (variância amostral) o estimador de σ^2.

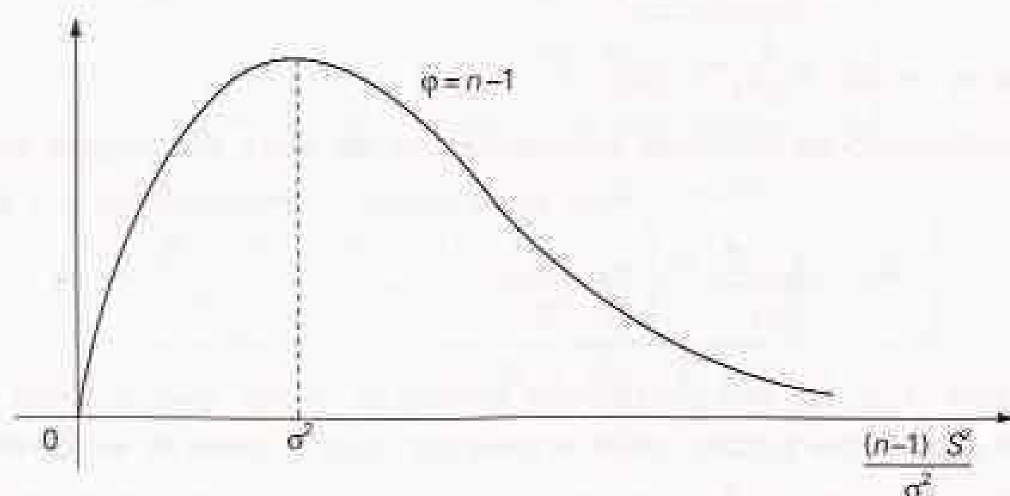
Se se desejar saber qual é a distribuição de S^2? Pode-se demonstrar que:

$$E[S^2] = \sigma^2 \quad \text{e} \quad \text{Var}[S^2] = \frac{2\sigma^4}{n-1}$$

e que S^2 tem distribuição qui-quadrado com $(n - 1)$ graus de liberdade. Ou seja:

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \overset{d}{=} \chi^2_{n-1}}$$

Lembre-se que $(n-1)$ e σ^2 são constantes. Graficamente, a relação entre S^2 e σ^2 é dada por uma distribuição qui-quadrado.



6.6 DISTRIBUIÇÃO AMOSTRAL DA SOMA, OU DIFERENÇA DE DUAS MÉDIAS

Deseja-se encontrar a distribuição amostral do estimador $(\bar{x}_1 \pm \bar{x}_2)$.

Sabe-se, conforme o item 6.4, que:

$$\bar{x}_1 \overset{d}{=} \left(\mu_1; \frac{\sigma_1^2}{n_1}\right) \quad \text{e} \quad \bar{x}_2 \overset{d}{=} \left(\mu_2; \frac{\sigma^2}{n_2}\right)$$

Considerando-se amostras independentes de duas populações tem-se:

$$\boxed{(\bar{x}_1 \pm \bar{x}_2) \overset{d}{=} N\left(\mu_1 \pm \mu_2 ; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}$$

Ou seja $(\bar{x}_1 \pm \bar{x}_2)$ tem distribuição normal de média igual à soma ou diferença das médias populacionais, e variância igual à soma das variâncias.

6.7 DISTRIBUIÇÃO AMOSTRAL DA SOMA, OU DIFERENÇA DE DUAS FREQÜÊNCIAS RELATIVAS

Deseja-se saber qual é a distribuição amostral de $(f_1 \pm f_2)$.

Sabe-se, conforme o item 6.5, que:

$$f_1 \overset{d}{=} N\left(p_1; \frac{p_1 \cdot q_1}{n_1}\right) \text{ e } f_2 \overset{d}{=} N\left(p_2; \frac{p_2 \cdot q_2}{n_2}\right)$$

Com $n_1 \geqslant 30$ e $n_2 \geqslant 30$.

Considerando-se amostras independentes de duas populações tem-se:

$$(f_1 \pm f_2) \overset{d}{=} N\left(p_1 \pm p_2; \frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}\right)$$

Ou seja $(f_1 \pm f_2)$ tem distribuição normal de média igual à soma ou diferença das proporções populacionais e variância igual à soma de suas variâncias.

6.8 DISTRIBUIÇÃO AMOSTRAL DAS MÉDIAS QUANDO A VARIÂNCIA DA POPULAÇÃO É DESCONHECIDA

Como foi visto, $\bar{x} \overset{d}{=} \left(\mu; \frac{\sigma^2}{n}\right)$. Ou seja, sua distribuição normal, padronizada é dada por:

$$Z_i = \frac{\bar{x}_i - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Como não se conhece o valor de σ^2, portanto de σ, uma possibilidade é substituir o σ pela variável aleatória S (desvio-padrão amostral) e procurar a distribuição da estatística T.

$$T = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

Pode-se demonstrar que T tem distribuição de Student com $n - 1$ graus de liberdade. Assim:

$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

Para o caso da soma ou diferença entre duas médias, admitindo-se variâncias desconhecidas e iguais tem-se:

$$t(n_1 + n_2 - 2) = \frac{(\overline{x}_1 \pm \overline{x}_2) - (\mu_1 - \mu_2)}{Sc\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

onde Sc é o desvio-padrão comum dado por:

$$Sc = \sqrt{\frac{(n_1 - 1)\,S_1^2 + (n_2 - 1)\,S_2^2}{n_1 + n_2 - 2}}$$

6.9 DISTRIBUIÇÃO AMOSTRAL DE RAZÕES DE VARIÂNCIAS

Deseja-se conhecer a distribuição amostral da razão: $\dfrac{S_1^2}{S_2^2}$.

Dadas duas amostras aleatórias independentes: n_1 e n_2, então a estatística:

$$F = \frac{\dfrac{n_1\,(n_1 - 1)\,S_1^2}{\sigma_1^2}}{\dfrac{n_2\,(n_2 - 1)\,S_2^2}{\sigma_2^2}}$$

tem distribuição F com $(n_1 - 1)$ e $(n_2 - 1)$ graus de liberdade. Ou seja, à exceção das constantes, S_1^2/S_2^2 tem distribuição F.

EXERCÍCIOS – SÉRIE I – CAPÍTULO 6

1. Uma população se constitui dos números 2, 3, 4, 5. Considere todas as amostras possíveis, de tamanho 2, que podem ser extraídas dessa população com reposição. Determine: a) a média da população, b) o desvio-padrão da população, c) a média da distribuição amostral das médias amostrais, d) o desvio-padrão da distribuição amostral das médias. Constate que:

$$\mu(\overline{x}) = \mu \quad \text{e} \quad \sigma(\overline{x}) = \frac{\sigma}{\sqrt{n}}$$

2. Considere os dados da população do exercício anterior e amostras de tamanho 2 sem reposição. Constate que:

$$\mu(\bar{x}) = \mu \quad \text{e} \quad \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} \sqrt{\left(\frac{N-n}{N-1}\right)}$$

3. Diz-se que um estimador $\hat{\theta}$ é justo, ou não tendencioso se $E[\hat{\theta}] = \theta$. Dê três exemplos de estimadores justos.

4. Uma caixa contém três peças boas e uma defeituosa. Considere amostras de tamanho 2, e sucesso à extração de uma peça boa.

   a) Supondo o processo com reposição, comprove os resultados:

$$\mu(f) = p \quad \text{e} \quad \sigma^2(f) = \frac{p \cdot q}{n}$$

   b) Supondo o processo sem reposição, comprove os resultados:

$$\mu(f) = p \ \text{ e } \ \sigma(f) = \frac{p \cdot q}{n}\left(\frac{N-n}{N-1}\right)$$

5. Uma amostra simples ao acaso de 30 domicílios foi selecionada de uma zona urbana que contém 15.000 domicílios. O número de pessoas de cada um dos domicílios que integram a amostra é o seguinte:

5	6	3	3	2	3	3	3	4	4	3	2	7	4	3
5	4	4	3	3	4	3	3	1	2	4	3	4	2	4

Estimar o número total de pessoas que vivem nesta área. Use o estimador $x' = N\bar{x}$.

6. As assinaturas de um requerimento foram colhidas em 676 folhas. Cada folha comporta 42 assinaturas, mas em muitas folhas foi registrado número menor de assinaturas. Retirou-se uma amostra de 50 folhas, obtendo-se

Nº de assinaturas	42	41	36	32	29	27	23	19	16	15	14	11	10	9	7	6	5	4	3
Nº de folhas	23	4	1	1	1	2	1	1	2	2	1	1	1	1	1	3	2	1	1

Estimar o número total de assinaturas.

7. Observe um jornal e dê três exemplos de estimativas.

7

Amostragem

7.1 INTRODUÇÃO

Geralmente, as pesquisas são realizadas através de estudo dos elementos que compõem uma amostra extraída da população que se pretende analisar. O conceito de população é intuitivo; trata-se do conjunto de indivíduos ou objetos que apresentam em comum determinadas características definidas para o estudo. Amostra é um subconjunto da população.

É compreensível que o estudo de todos os elementos da população possibilita preciso conhecimento das variáveis que estão sendo pesquisadas; todavia, nem sempre é possível obter as informações de todos os elementos da população. Limitações de tempo, custo e as vantagens do uso das técnicas estatísticas de inferências justificam o uso de planos amostrais. Torna-se claro que a representatividade da amostra dependerá do seu tamanho (quanto maior, melhor) e de outras considerações de ordem metodológica. Isto é, o investigador procurará acercar-se de cuidados, visando à obtenção de uma amostra significativa, ou seja, que de fato represente "o melhor possível" toda a população.

Na teoria da amostragem, são consideradas duas dimensões:

a) Dimensionamento da amostra.

b) Composição da amostra.

7.2 DIMENSIONAMENTO DA AMOSTRA

Procedimento:

1º) Analise o questionário, ou roteiro da entrevista e escolha uma variável que julgue mais importante para o estudo. Se possível, escolha mais do que uma.

2º) Verifique o nível de mensuração da variável: se nominal, ordinal ou intervalar.

3º) Considere o tamanho da população: infinita ou finita.

4º) Se a variável escolhida for intervalar e a população considerada infinita, você poderá determinar o tamanho da amostra pela fórmula:

$$n = \left(\frac{Z \cdot \sigma}{d} \right)^2$$

onde: Z = abscissa da curva normal padrão, fixado um nível de confiança.
Se o nível for 95,5%, $Z = 2$
Se o nível for 95%, $Z = 1,96$
Se o nível for 99%, $Z = 2,57$

Geralmente, utiliza-se $Z = 2$.

σ = desvio-padrão da população, expresso na unidade variável. Você poderá determiná-lo de pelo menos três maneiras:
- Especificações técnicas
- Resgatar o valor de estudos semelhantes
- Fazer conjeturas sobre possíveis valores.

d = erro amostral, expresso na unidade da variável. O erro amostral é a máxima diferença que o investigador admite suportar entre μ e $\bar{x}$, isto é: $|\mu - \bar{x}| < d$, onde μ é a verdadeira média populacional, que ele não conhece, e $\bar{x}$ será a média amostral a ser calculada a partir da amostra.

5º) Se a variável escolhida for intervalar e a população finita, tem-se:

$$n = \frac{Z^2 \cdot \sigma^2 \cdot N}{d^2 (N - 1) + Z^2 \sigma^2}$$

onde: Z = abscissa da normal padrão (veja comentário anterior)

σ = desvio-padrão da população (veja comentário anterior)

N = tamanho da população

d = erro amostral (veja comentário anterior)

6º) Se a variável escolhida for nominal ou ordinal, e a população considerada infinita, você poderá determinar o tamanho da amostra pela fórmula:

$$n = \frac{Z^2 \cdot \hat{p} \cdot \hat{q}}{d^2}$$

onde: Z = abscissa da normal padrão (veja comentário do item 4º)

$\hat{p}$ = estimativa da verdadeira proporção de um dos níveis da variável escolhida. Por exemplo, se a variável escolhida for porte da empresa, $\hat{p}$ poderá ser a estimativa da verdadeira proporção de grandes empresas do setor que está sendo estudado. Será expresso em decimais. Assim, se $\hat{p} = 30\%$, teremos: $\hat{p} = 0{,}30$.

$\hat{q} = 1 - \hat{p}$

d = erro amostral, expresso em decimais. O erro amostral neste caso será a máxima diferença que o investigador admite suportar entre p e $\hat{p}$, isto é: $|p - \hat{p}| < d$, em que p é a verdadeira proporção, que ele não conhece, e $\hat{p}$ será a proporção (freqüência relativa) do evento a ser calculado a partir da amostra.

7º) Se a variável escolhida for nominal ou ordinal e a população finita, tem-se:

$$n = \frac{Z^2 \cdot \hat{p} \cdot \hat{q} \cdot N}{d^2 (N - 1) + Z^2 \cdot \hat{p} \cdot \hat{q}}$$

onde: N = tamanho da população

Z = abscissa da normal padrão (veja item 4º)

$\hat{p}$ = estimativa da proporção (veja item 6º)

$\hat{q}$ = $1 - p$ (veja item 6º)

d = erro amostral (veja item 6º)

Estas fórmulas são básicas para qualquer tipo de composição da amostra; todavia, existem fórmulas específicas segundo o critério de composição da amostra.

Se o investigador escolhe mais de uma variável, deve optar pelo maior "n" obtido.

Eis alguns exemplos:

I – Suponha que a variável escolhida num estudo seja o peso de certa peça e que a população é infinita. Pelas especificações do produto, o desvio-padrão (dispersão em torno da média) é de 10 kg. Logo, admitindo-se um nível de confiança de 95,5% e um erro amostral de 1,5 kg, tem-se:

$$n = \left(\frac{2 \times 10}{1{,}5}\right)^2 = 177{,}77 \cong 178$$

II – Admita os mesmos dados do exemplo anterior e que a população seja finita de 600 peças. Logo:

$$n = \frac{2^2 \cdot 10^2 \cdot 600}{1{,}5^2\,(600 - 1) + 2^2 \cdot 10^2} = 137{,}31 \cong 138$$

III – Suponha que a variável escolhida num estudo seja a proporção de eleitores favoráveis ao candidato *X* e que o investigador tenha elementos para suspeitar que essa porcentagem seja de 30%. Admita a população infinita e que se deseja um nível de confiança de 99% e um erro amostral de 2% (ou seja: que a diferença entre a verdadeira proporção de eleitores do candidato *X* e a estimativa a ser calculada na amostra seja no máximo de 2%). Assim:

$$Z = 2{,}57 \qquad \hat{p} = 30\% = 0{,}30 \qquad \hat{q} = 1 - 0{,}30 = 0{,}70$$

$$d = 2\% = 0{,}02 \qquad n = \frac{(2{,}57)^2\,(0{,}30)\,(0{,}70)}{(0{,}02)^2} = 3.467{,}57 \cong 3468$$

IV – Admita os mesmos dados do exemplo anterior, e que a população de eleitores seja finita de 20.000 eleitores. Logo:

$$n = \frac{(2{,}57)^2 \cdot (0{,}30)\,(0{,}70)\,(20.000)}{(0{,}02)^2\,(20.000 - 1) + (2{,}57)^2\,(0{,}30)\,(0{,}70)} = 2955{,}33 \cong 2956$$

Observação: Quando você não tiver condições de prever o valor de "$\hat{p}$", admita $\hat{p} = 0{,}50$, pois, dessa forma, você terá o maior tamanho da amostra, admitindo-se constantes os demais elementos.

7.3 COMPOSIÇÃO DA AMOSTRA

Basicamente, existem dois métodos para composição da amostra: probabilístico e não probabilístico ou intencional.

a) *Métodos Probabilísticos*

O método de amostragem probabilística exige que cada elemento da população possua determinada probabilidade de ser selecionado. Normalmente possuem a mesma probabilidade. Assim, se N for o tamanho da população, a probabilidade de cada elemento será $1/N$. Trata-se do método que garante cientificamente a aplicação das técnicas estatísticas de inferências. Somente com base em amostragens probabilísticas é que se podem realizar inferências ou induções sobre a população a partir do conhecimento da amostra.

7.3.1 Amostragem Aleatória Simples

É o processo mais elementar e freqüentemente utilizado. Atribui-se a cada elemento da população um número distinto. Se a população for numerada utilizam-se esses "rótulos". Efetuam-se sucessivos sorteios até completar-se o tamanho da amostra: n. Para realizar os sorteios, utilizam-se "tábuas de números aleatórios" que consistem em tabelas que apresentam seqüências dos dígitos de 0 a 9 distribuídos aleatoriamente.

Se, por exemplo, a população tem 1.000 elementos (N = 1.000), pode-se numerá-los de 000 a 999. Escolhendo uma posição de qualquer linha da tabela de números aleatórios, faz-se o sorteio, ou seja, retiram-se conjuntos de três algarismos para se escolherem os elementos que irão compor a amostra. Assim, imagine-se que a seqüência de dígitos aleatórios seja: 385 559 555 432 886 ... logo, os elementos de números 385 – 559 – 555 – 432 ... serão os componentes da amostra.

Se o número sorteado superar o maior número dos elementos rotulados, abandona-se o número sorteado, prosseguindo-se o processo. Se o número sorteado for repetido, convém abandoná-lo.

No final do livro, você encontrará uma tabela de dígitos aleatórios.

7.3.2 Amostragem Sistemática

Trata-se de uma variação da amostragem aleatória simples, conveniente quando a população está ordenada segundo algum critério, como fichas em um fichário, listas telefônicas ...

Calcula-se o intervalo de amostragem N/n aproximando-o para o inteiro mais próximo: a. Utilizando-se a tábua dos números aleatórios, sorteia-se um número x entre 1 e a, formando-se a amostra dos elementos correspondentes aos números x; $x + a$; $x + 2a$; ...

Por exemplo, seja $N = 1.000$, $n = 200$. Logo:

$$a = \frac{N}{n} = \frac{1.000}{200} = 5.$$

Imagine que três seja o número sorteado entre 1 e 5. Portanto, os elementos da população numerados por 3, 8, 13, ..., 998 irão compor a amostra.

7.3.3 Amostragem Estratificada

No caso de população heterogênea em que se podem distinguir subpopulações mais ou menos homogêneas, denominadas estratos, é possível utilizar o processo de amostragem estratificada.

Após a determinação dos estratos, seleciona-se uma amostra aleatória de cada subpopulação (estrato).

Se as diversas subamostras tiverem tamanhos proporcionais aos respectivos números de elementos dos estratos, e guardarem proporcionalidade com respeito à variabilidade de cada estrato, obtém-se uma estratificação ótima.

As variáveis de estratificação mais comuns são: classe social, idade, sexo, profissão ... ou qualquer outro atributo que revele os estratos dentro da população.

7.3.4 Amostragem por Conglomerados (ou Agrupamentos)

Algumas populações não permitem, ou tornam extremamente difícil que se identifiquem seus elementos. Não obstante isso, pode ser relativamente fácil

identificar alguns subgrupos da população. Em tais casos, uma amostra aleatória simples desses subgrupos (conglomerados) pode ser colhida, e uma contagem completa deve ser feita para o conglomerado sorteado. Agregados típicos são quarteirões, famílias, organizações, agências, edifícios etc.

Assim, por exemplo, num levantamento da população de uma cidade, podemos dispor do mapa indicando cada quarteirão e não dispor de uma relação atualizada dos seus moradores. Pode-se, então, colher uma amostra dos quarteirões e fazer a contagem completa de todos os que residem naqueles quarteirões sorteados.

b) *Métodos Não Probabilísticos*

São amostragens em que há uma escolha deliberada dos elementos da amostra. Não é possível generalizar os resultados das pesquisas para a população, pois as amostras não-probabilísticas não garantem a representatividade da população.

7.3.5 Amostragem Acidental

Trata-se de uma amostra formada por aqueles elementos que vão aparecendo, que são possíveis de se obter até completar o número de elementos da amostra. Geralmente utilizada em pesquisas de opinião, em que os entrevistados são acidentalmente escolhidos.

7.3.6 Amostragem Intencional

De acordo com determinado critério, é escolhido intencionalmente um grupo de elementos que irão compor a amostra. O investigador se dirige intencionalmente a grupos de elementos dos quais deseja saber a opinião. Por exemplo, numa pesquisa sobre preferência por determinado cosmético, o pesquisador se dirige a um grande salão de beleza e entrevista as pessoas que ali se encontram.

7.3.7 Amostragem por Quotas

Um dos métodos de amostragem mais comumente usados em levantamentos de mercado e em prévias eleitorais é o método de amostragem por quotas. Ele abrange três fases:

1) classificação da população em termos de propriedades que se sabe, ou presume, serem relevantes para a característica a ser estudada;

2) determinação da proporção da população para cada característica, com base na constituição conhecida, presumida ou estimada, da população; e

3) fixação de quotas para cada observador ou entrevistador a quem tocará a responsabilidade de selecionar interlocutores ou entrevistados, de modo que a amostra total observada ou entrevistada contenha a proporção de cada classe tal como determinada em (2).

Exemplificando: Admite-se que se deseja pesquisar o "trabalho das mulheres". Provavelmente se terá interesse em considerar: a divisão cidade/campo, a habitação, o número de filhos, a idade dos filhos, a renda média, as faixas etárias ...

A primeira tarefa é descobrir as proporções (porcentagens) dessas características na população. Imagine-se que haja 47% de homens e 53% de mulheres na população. Logo, uma amostra de 50 pessoas deverá ter 23 homens e 27 mulheres. Então o pesquisador receberá uma "quota" para entrevistar 27 mulheres. A consideração de várias categorias exigirá uma composição amostral que atenda ao *n* determinado e às proporções populacionais estipuladas.

EXERCÍCIOS – SÉRIE I

1. Dada a seguinte população (rendas em $ 1000)

29	6	34	12	15	31	34	20	8	30
8	15	24	22	35	31	25	26	20	10
30	4	16	21	14	21	16	18	20	12
31	20	12	18	12	25	26	13	10	5
13	19	30	17	25	29	25	28	32	15
10	21	18	7	16	14	11	22	21	36
32	17	15	13	8	12	23	25	13	21
5	12	32	21	10	30	30	10	14	17
34	22	30	48	19	12	8	7	15	20
26	25	22	30	33	14	17	13	10	9

a) calcule o tamanho da amostra para se estimar a média, sendo d = $ 2000, σ = $ 7000 e $1 - \alpha = 95{,}5\%$;

b) retire uma amostra aleatória simples considerando o tamanho amostral obtido em a);

c) agrupe os elementos da amostra em classes;

d) calcule sua média;

e) calcule o desvio-padrão amostral;

f) calcule a média da população e verifique se $|\mu - \bar{x}| \leq d$.

2. Escolha uma página qualquer da lista telefônica e retire uma amostra sistemática de cinqüenta nomes.

3. Calcule o tamanho da amostra de seus colegas desta faculdade, para estimar a proporção dos usuários de óculos.

4. Sendo $\hat{p} = \hat{q} = 0{,}5$, população infinita, $d = 0{,}05$ e $1 - \alpha = 95{,}5\%$, determine o tamanho amostral.

5. Sendo $\hat{p} = \hat{q} = 0{,}5$, população de 200.000, $d = 0{,}05$ e $1 - \alpha = 95{,}5\%$, determine o tamanho amostral. Compare com o resultado obtido no Exercício 4.

6. Determine o tamanho amostral para se estimar o salário médio dos trabalhadores do município em que você mora.

8

INTERVALO DE CONFIANÇA

8.1 INTRODUÇÃO

Trata-se de uma técnica para se fazer inferência estatística. Ou seja, a partir de um intervalo de confiança, construído com os elementos amostrais, pode-se inferir sobre um parâmetro populacional.

A construção de intervalos de confiança fundamenta-se nas distribuições amostrais vistas no Capítulo 6. Sua "lógica" é a seguinte:

Seja θ um parâmetro populacional.

Seja $\hat{\theta}$ um estimador de θ.

Conhecida a distribuição de probabilidade de $\hat{\theta}$, é possível construir um intervalo:

$$\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2$$

que contém θ, e se exigir que a probabilidade do intervalo seja de $(1 - \alpha)$ = nível de confiança.

Geralmente $(1 - \alpha)\,100 = 90\%; 95\%; 99\% \ldots$

Esta técnica diferencia-se da estimação "por ponto", onde se calcula um único valor (estimativa) para o parâmetro populacional. No caso do intervalo de confiança busca-se um "segmento", ou intervalo $\hat{\theta}_1$; $\hat{\theta}_2$ que contém o parâmetro desconhecido.

Por exemplo, retira-se uma amostra de 500 brasileiros e calcula-se a média de suas alturas encontrando-se 1,66 m. Logo, uma estimação pontual da verdadeira altura média (μ) é dada por $\bar{x}$ = 1,66 m. Já através do intervalo de confiança poder-se-ia encontrar um intervalo, por exemplo [1,58; 1,68] que, em 95% das vezes, incluiria μ (a verdadeira altura média dos brasileiros).

Os cálculos das medidas de posição e dispersão vistos no Capítulo 5 constituem exemplos de estimações pontuais.

Neste capítulo serão apresentadas as estimativas por intervalos.

8.2 INTERVALO DE CONFIANÇA PARA A MÉDIA POPULACIONAL (μ) QUANDO A VARIÂNCIA (σ^2) É CONHECIDA

Como se sabe, o estimador de μ é $\bar{x}$. Também é conhecida a distribuição de probabilidade de $\bar{x}$:

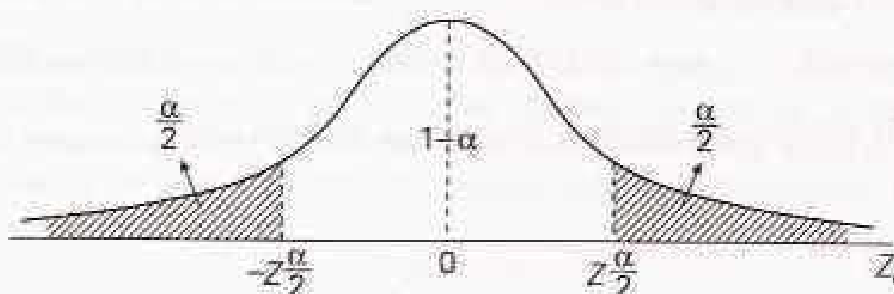
$\bar{x} \overset{d}{=} N\left(\mu; \dfrac{\sigma^2}{n}\right)$ para as populações infinitas.

$\bar{x} \overset{d}{=} N\left(\mu; \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)\right)$ para as populações finitas.

Assim, para o caso de populações infinitas, a variável padronizada de $\bar{x}$ será:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Fixando-se um nível de confiança: $1 - \alpha$ tem-se:



Ou seja: $P\left(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$

Substituindo-se o valor de Z,

$$P\left(-Z\frac{\alpha}{2} \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq Z\frac{\alpha}{2}\right) = 1-\alpha$$

Resolvendo-se as duas inequações para μ, tem-se o intervalo de confiança para a média populacional (μ) quando a variância (σ^2) é conhecida:

$$P\left(\bar{x} - Z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

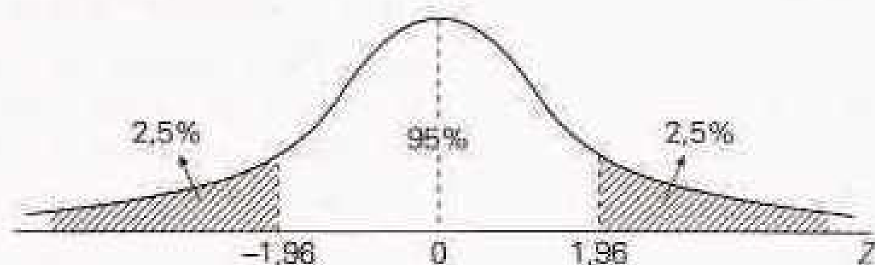
Como poderá ser verificado, a aplicação da fórmula é extremamente simples. Fixa-se o valor de $1-\alpha$, ou $(1-\alpha)\,100 = \%$, observa-se na tabela da distribuição normal padrão o valor das abscissas que deixam $\frac{\alpha}{2}$ em cada uma das caudas. Com os valores de $\bar{x}$ (média amostral), σ = desvio-padrão da população, que neste caso é conhecido, e n (tamanho da amostra), constrói-se o intervalo.

Exemplo: A duração da vida de uma peça de equipamento é tal que $\sigma = 5$ horas. Foram amostradas 100 dessas peças obtendo-se a média de 500 horas. Deseja-se construir um intervalo de confiança para a verdadeira duração média da peça com um nível de 95%.

Solução: Do problema se tem:

$\sigma = 5$; $n = 100$ $\bar{x} = 500$ $(1-\alpha)\,100 = 95\%$

O gráfico da distribuição normal padrão será:



Lembre-se que para descobrir a abscissa 1,96, entrou-se na tabela com 0,475 = 47,5%, já que a tabela é de faixa central.

Substituindo-se os dados na fórmula:

$$P\left(500 - 1{,}96 \cdot \frac{5}{\sqrt{100}} \leq \mu \leq 500 + 1{,}96 \cdot \frac{5}{\sqrt{100}}\right) = 95\%$$

Efetuando-se os cálculos:

$$P(499{,}02 \leq \mu \leq 500{,}98) = 95\%$$

que é o intervalo solicitado.

A interpretação desse resultado é dada por:

O intervalo [499,02; 500,98] contém a duração média da peça com 95% de confiança.

Isto significa que se forem construídos intervalos dessa mesma maneira, para um grande número de amostras, em 95% dos casos tais intervalos incluiriam μ.

Para o caso de populações finitas usa-se a seguinte fórmula:

$$P\left(\bar{x} - Z\frac{\alpha}{2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{x} + Z\frac{\alpha}{2} \cdot \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

8.3 INTERVALO DE CONFIANÇA PARA A MÉDIA (μ) QUANDO A VARIÂNCIA (σ^2) É DESCONHECIDA

O processo para se obter o intervalo de confiança é semelhante àquele mostrado no item anterior. Como não se conhece σ, porém, é preciso substituí-lo por S (desvio-padrão amostral) que, contrariamente a σ, é uma variável aleatória. Daí se ter o quociente entre duas variáveis aleatórias, $\bar{x}$ e S, pois:

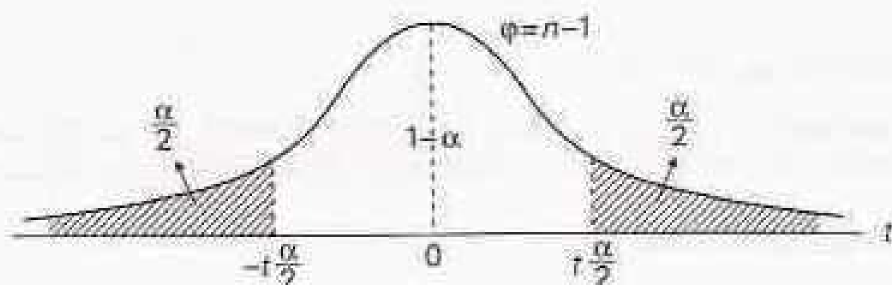
$$\frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

Pode-se demonstrar que:

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

tem distribuição "t" de Student com $(n-1)$ graus de liberdade.

Fixando-se um nível de confiança: $1-\alpha$ tem-se:



Ou seja: $P\left(-t\frac{\alpha}{2} \leq t \leq t\frac{\alpha}{2}\right) = 1-\alpha$

Substituindo-se o valor de "t" e resolvendo-se as inequações para μ ob-tém-se o intervalo para a média quando a variância (σ^2) é desconhecida.

$$P\left(\bar{x} - t\frac{\alpha}{2} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t\frac{\alpha}{2} \cdot \frac{S}{\sqrt{n}}\right) = 1-\alpha$$

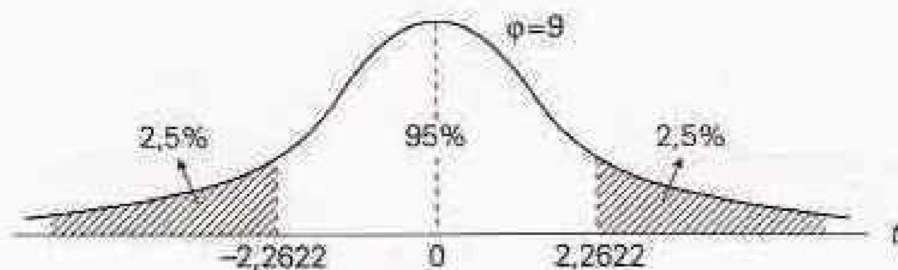
onde a variável "t" possui $(n-1)$ graus de liberdade.

Exemplo: A amostra: 9, 8, 12, 7, 9, 6, 11, 6, 10, 9 foi extraída de uma população normal. Construir um intervalo de confiança para a média ao nível de 95%.

Solução: Calculando-se a média e o desvio-padrão da amostra obtém-se:

$\bar{x} = 8{,}7$ e $S = 2$

Como: $1-\alpha = 95\%$ e g.l. = 9 pois g.l. = $\varphi = n-1 = 10-1 = 9$



Confira os valores de $t_{0{,}025}$ na tabela.

Logo:

$$P\left(8{,}7 - 2{,}2622\ \frac{2}{\sqrt{10}} \leq \mu \leq 8{,}7 + 2{,}2622 \cdot \frac{2}{\sqrt{10}}\right) = 95\%$$

ou:

$$P(7{,}27 \leq \mu \leq 10{,}13) = 95\%$$

A interpretação desse resultado é dada por:

"O intervalo [7,27; 10,13] contém a verdadeira média com 95% de confiança."

Para o caso de populações finitas usa-se a seguinte fórmula:

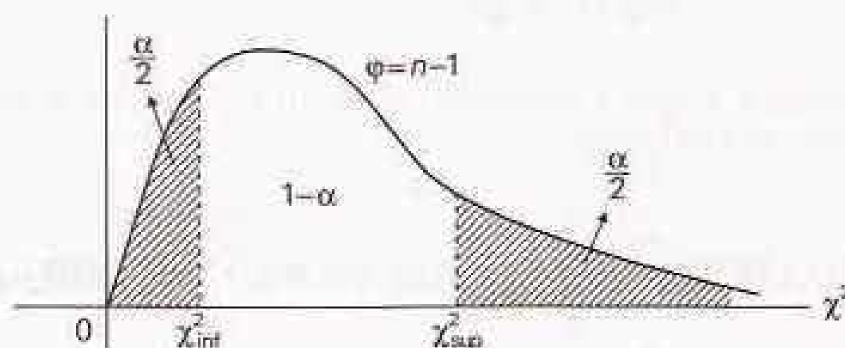
$$P\left(\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \leq \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

8.4 INTERVALO DE CONFIANÇA PARA A VARIÂNCIA

O estimador de σ^2 e S^2. Demonstra-se que $\frac{(n-1) \cdot S^2}{\sigma^2}$ tem distribuição qui-quadrado com $(n-1)$ graus de liberdade. Ou seja:

$$\chi_{n-1} \stackrel{d}{=} \frac{(n-1)\,S^2}{\sigma^2}$$

Então, o intervalo será:



Substituindo-se o valor de χ^2, e isolando-se σ^2 obtém-se:

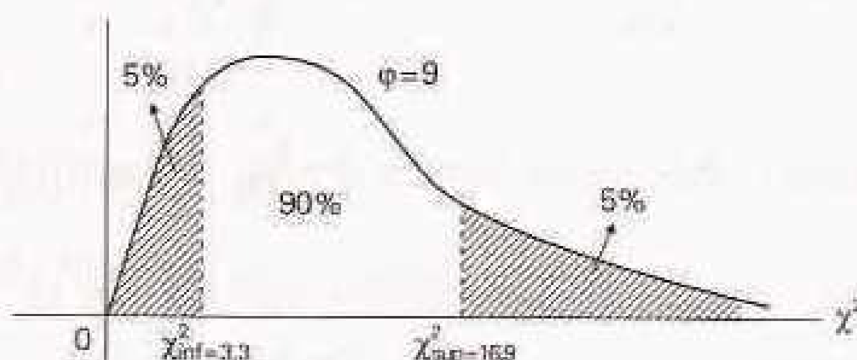
$$P\left(\frac{(n-1)\cdot S^2}{\chi^2_{sup}} \leq \sigma^2 \leq \frac{(n-1)\cdot S^2}{\chi^2_{inf}}\right) = 1 - \alpha$$

que é o IC (intervalo de confiança) para σ^2. Onde a distribuição qui-quadrado possui parâmetro: $(n - 1)$.

Exemplo: Admita $n = 10$, $S^2 = 4$ e que se deseja construir um IC para a variância ao nível 90%.

Solução: tem-se: $n = 10$, $S^2 = 4$, $1 - \alpha = 90\%$ e $\varphi = (n - 1) = (10 - 1) = 9$

Consultando-se a tabela da distribuição qui-quadrado:



Logo:

$$P\left(\frac{9(4)}{1,69} \leq \sigma^2 \leq \frac{9(4)}{3,33}\right) = 90\%$$

$$P(2,13 \leq \sigma^2 \leq 10,81) = 90\%$$

A interpretação é que o intervalo [2,13; 10,81] contém a verdadeira variância com 90% de confiança.

8.5 INTERVALO DE CONFIANÇA PARA O DESVIO-PADRÃO

Como o desvio-padrão é a raiz quadrada da variância, pode-se usar a seguinte fórmula:

$$P\left(S \cdot \sqrt{\frac{(n-1)}{\chi^2_{sup}}} \leq \sigma \leq S \cdot \sqrt{\frac{(n-1)}{\chi^2_{inf}}}\right) = 1 - \alpha$$

Com a distribuição qui-quadrado de parâmetro: $\varphi = (n - 1)$.

A interpretação segue o modelo já apresentado.

8.6 INTERVALO DE CONFIANÇA PARA PROPORÇÃO; OU PROBABILIDADE (P)

Foi visto no capítulo anterior que f, o estimador de "p", tem distribuição dada por:

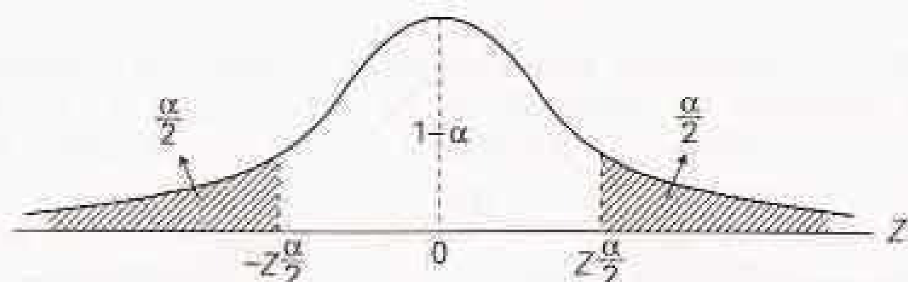
$f \stackrel{d}{=} N\left(p; \frac{p \cdot q}{n}\right)$ para população infinita e,

$f \stackrel{d}{=} N\left(p; \frac{p \cdot q}{n}\left(\frac{N-n}{N-1}\right)\right)$ para população finita.

Assim, para o caso de populações infinitas, a variável padronizada de f é dada por:

$$Z = \frac{f - p}{\sqrt{\frac{p \cdot q}{n}}}$$

Fixando-se um nível de confiança $1 - \alpha$ tem-se:



Ou seja: $P\left(-Z\frac{\alpha}{2} \leq Z \leq Z\frac{\alpha}{2}\right) = 1 - \alpha$

Substituindo-se o valor de Z:

$$P\left(-Z\frac{\alpha}{2} \leq \frac{f - p}{\sqrt{\frac{p \cdot q}{n}}} \leq Z\frac{\alpha}{2}\right) = 1 - \alpha$$

Isolando-se "p" do denominador, encontra-se:

$$P\left(f - Z\frac{\alpha}{2}\sqrt{\frac{p \cdot q}{n}} \leq p \leq f + Z\frac{\alpha}{2}\sqrt{\frac{p \cdot q}{n}}\right) = 1 - \alpha$$

Para amostras grandes ($n > 30$) pode-se substituir "p" e $q = 1 - p$ do radicando por "f" e $(1 - f)$. Assim, o IC para a proporção será:

$$\boxed{P\left(f - Z\frac{\alpha}{2}\sqrt{\frac{f(1 - f)}{n}} \leq p \leq f + Z\frac{\alpha}{2}\sqrt{\frac{f(1 - f)}{n}}\right) = 1 - \alpha}$$

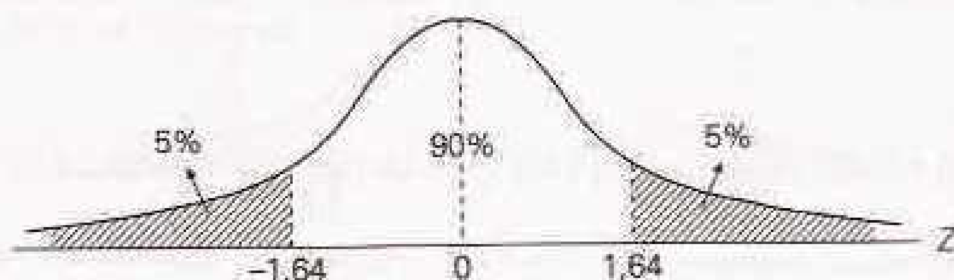
Para o caso de populações finitas o IC será:

$$\boxed{P\left(f - Z\frac{\alpha}{2}\sqrt{\frac{f(1-f)}{n}\left(\frac{N-n}{N-1}\right)} \leq p \leq f + Z\frac{\alpha}{2}\sqrt{\frac{f(1-f)}{n}\left(\frac{N-n}{N-1}\right)}\right) = 1 - \alpha}$$

Exemplo: Examinadas 500 peças de uma grande produção encontrou-se 260 defeituosas. No nível de 90% construir um IC para a verdadeira proporção de peças defeituosas.

Solução: Tem-se: $n = 500$, $x = 260$, $1 - \alpha = 90\%$

Logo: $f = \frac{x}{n} = \frac{260}{500} = 0{,}52$



Então, o IC será:

$$P\left(0{,}52 - 1{,}64\sqrt{\frac{0{,}52\,(1-0{,}52)}{500}} \leqslant p \leqslant 0{,}52 + 1{,}64\sqrt{\frac{0{,}52\,(1-0{,}52)}{500}}\right) = 90\%$$

ou: $P(0{,}488 \leqslant p \leqslant 0{,}552) = 90\%$, ou ainda:

$P(48{,}8\% \leqslant p \leqslant 55{,}2\%) = 90\%$,

E a interpretação é de que o intervalo [48,8%; 55,2%] contém a verdadeira porcentagem (ou proporção) de peças defeituosas.

EXERCÍCIOS – SÉRIE I – CAPÍTULO 8
Construa os IC e interprete cada resultado

Para a Média Populacional

1. Foram retiradas 25 peças da produção diária de uma máquina, encontrando-se para uma medida uma média de 5,2 mm. Sabendo-se que as medidas têm distribuição normal com desvio-padrão populacional 1,2 mm, construir intervalos de confiança para a média aos níveis de 90%, 95% e 99%.

2. De uma distribuição normal com $\sigma^2 = 1{,}96$, obteve-se a seguinte amostra: 25,2; 26,0; 26,4; 27,1; 28,2; 28,4. Determinar o intervalo de confiança para a média da população, sendo $\alpha = 0{,}05$ e $\alpha = 0{,}10$.

3. Suponha que as alturas dos alunos de nossa faculdade tenham distribuição normal com $\sigma = 15$ cm. Foi retirada uma amostra aleatória de 100 alunos obtendo-se $\bar{x} = 175$ cm. Construir, ao nível de significância de 95% o intervalo para a verdadeira altura média dos alunos.

4. Dados n = 10, $\bar{x} = 110$ e $S = 10$, determinar os intervalos de confiança μ aos níveis de 90% e 95%.

5. Uma amostra é composta pelos seguintes elementos: 7, 7, 8, 9, 9, 9, 10, 11, 11, 11, 12, 13, 13, 14, 15, 15. Construir os intervalos de confiança para a média sendo: $1 - \alpha = 97{,}5\%$ e $1 - \alpha = 75\%$.

6. Colhida uma amostra de 30 peças, forneceu os seguintes pesos:

250,	265,	267,	269,	271,	275,	277,	281,	283,	284,
287,	289,	291,	293,	293,	298,	301,	303,	306,	307,
307,	309,	311,	315,	319,	322,	324,	328,	335,	339.

Por meio da construção do intervalo de confiança, responder se esta amostra satisfaz a especificação pela qual o peso médio deve ser 300 kg.

Sugestão: Adote $\alpha = 5\%$.

7. Em uma fábrica, colhida uma amostra de certa peça, obtiveram-se as seguintes medidas para os diâmetros:

10,	11,	11,	11,	12,	12,	12,	12,	13,	13,
13,	13,	13,	13,	13,	13,	13,	13,	13,	13,
14,	14,	14,	14,	14,	15,	15,	15,	16,	16.

a) Estimar a média e a variância;

b) Construir um intervalo de confiança para a média sendo $\alpha = 5\%$.

8. Em quatro leituras experimentais de um "comercial" de 30 segundos, um locutor levou em média 29,2 segundos com uma $S^2 = 5{,}76$ segundos ao quadrado. Construir os limites de confiança para a média. Dado $\alpha = 10\%$.

9. Construir intervalos de confiança para a média admitindo-se as seguintes distribuições amostrais, ao nível de 95%:

a)

Classes	0 – 5	5 – 10	10 – 15	15 – 20
F_i	2	3	5	2

b)

Classes	15 – 18	18 – 21	21 – 24	24 – 27	27 – 30	30 – 33
F_i	8	9	12	15	7	4

c)

Classes	2,2 – 6,2	6,2 – 10,2	10,2 – 14,2	14,2 – 18,2
F_i	3	4	5	3

Intervalos de Confiança para a Variância

10. Supondo populações normais, construir o intervalo de confiança para a variância ao nível de 90% para as amostras:

a) 44,9 – 44,1 – 43 – 42,9 – 43,2 – 44,5;

b) 2 – 2 – 2 – 3 – 3 – 4 – 5 – 5– 5 – 5 – 6 – 6 – 7 – 7 – 8.

11. Suponhamos que uma amostra de $n = 10$ fornecesse $S^2 = 2{,}25$. Quais os limites de confiança a 80% para a verdadeira variância?

12. Sendo X uma população tal que $X = N(\mu, \sigma^2)$ em que μ e σ^2 são desconhecidos. Uma amostra de tamanho 15 forneceu os valores $\Sigma x_i = 8{,}7$ e $\Sigma x_i^2 = 27{,}3$. Determinar um intervalo de confiança de 95% para σ^2.

13. Determinar, ao nível de 99%, o intervalo para o desvio-padrão da população que deu origem à amostra do exercício 6 desta série.

14. Qual é o intervalo de confiança que conterá com 90% a verdadeira variância de uma população normal que resultou $\Sigma X_i = 700{,}8$ e $\Sigma X_i^2 = 23.436{,}80$ de uma amostra de 30 elementos?

Intervalos de Confiança para a Proporção

15. Uma centena de componentes foi ensaiada e 93 deles funcionaram mais de 500 horas. Determinar um intervalo de confiança de 95% para a proporção.

16. Uma amostra aleatória de 400 domicílios mostra-nos que 25% deles são casas de aluguel. Qual é o intervalo de confiança da proporção de casas de aluguel? $\alpha = 2\%$.

17. Em 50 lances de uma moeda, foram obtidas 30 caras. A partir de um intervalo de confiança de 96%, pode-se dizer que a moeda é honesta?

18. Para verificar se um dado era viciado, jogou-se o mesmo 120 vezes, obtendo-se 25 vezes o número cinco. Calcular um intervalo de confiança para a proporção $\alpha = 1\%$. Pode-se dizer que o dado é viciado?

19. Uma amostra de 300 habitantes de uma cidade mostrou que 180 desejavam a água fluorada. Encontrar os limites de confiança de 90% e 95% para a proporção da população favorável a fluoração.

9

TESTE DE HIPÓTESES

9.1 INTRODUÇÃO

Trata-se de uma técnica para se fazer inferência estatística. Ou seja, a partir de um teste de hipóteses, realizado com os dados amostrais, pode-se inferir sobre a população.

No caso das inferências através da IC, busca-se "cercar" o parâmetro populacional desconhecido. Aqui formula-se uma hipótese quanto ao valor do parâmetro populacional, e pelos elementos amostrais faz-se um teste que indicará a aceitação ou rejeição da hipótese formulada.

9.2 PRINCIPAIS CONCEITOS

9.2.1 Hipótese Estatística

Trata-se de uma suposição quanto ao valor de um parâmetro populacional, ou quanto à natureza da distribuição de probabilidade de uma variável populacional.

Neste capítulo serão apresentados os testes referentes aos parâmetros da população.

São exemplos de hipóteses estatísticas:

a) A altura média da população brasileira é 1,65 m, isto é: $H: \mu = 1{,}65$ m

b) A variância populacional dos salários vale \$ 5.000^2, isto é $H: \sigma^2 = 5.000^2$.

9

TESTE DE HIPÓTESES

9.1 INTRODUÇÃO

Trata-se de uma técnica para se fazer inferência estatística. Ou seja, a partir de um teste de hipóteses, realizado com os dados amostrais, pode-se inferir sobre a população.

No caso das inferências através da IC, busca-se "cercar" o parâmetro populacional desconhecido. Aqui formula-se uma hipótese quanto ao valor do parâmetro populacional, e pelos elementos amostrais faz-se um teste que indicará a aceitação ou rejeição da hipótese formulada.

9.2 PRINCIPAIS CONCEITOS

9.2.1 Hipótese Estatística

Trata-se de uma suposição quanto ao valor de um parâmetro populacional, ou quanto à natureza da distribuição de probabilidade de uma variável populacional.

Neste capítulo serão apresentados os testes referentes aos parâmetros da população.

São exemplos de hipóteses estatísticas:

a) A altura média da população brasileira é 1,65 m, isto é: $H: \mu = 1{,}65$ m

b) A variância populacional dos salários vale \$ 5.000^2, isto é: $H: \sigma^2 = 5.000^2$.

c) A proporção de paulistas com a doença X é 40%, ou seja: H: $p = 0{,}40$.

d) A distribuição dos pesos dos alunos da nossa faculdade é normal.

e) A chegada de navios ao porto de Santos é descrita por uma distribuição Poisson.

9.2.2 Teste de Hipótese

É uma regra de decisão para aceitar ou rejeitar uma hipótese estatística com base nos elementos amostrais.

9.2.3 Tipos de Hipótese

Designa-se por H_0, chamada hipótese nula, a hipótese estatística a ser testada, e por H_1 a hipótese alternativa. A hipótese nula expressa uma igualdade, enquanto a hipótese alternativa é dada por uma desigualdade.

Exemplos:

a) H_0: $\mu = 1{,}65$ m
H_1: $\mu \neq 1{,}65$ m

Dará origem a um teste bicaudal

b) H_0: $\mu = 1{,}65$ m
H_1: $\mu > 1{,}65$ m

Dará origem a um teste unicaudal à direta

c) H_0: $\mu = 1{,}65$ m
H_1: $\mu < 1{,}65$ m

Dará origem a um teste unicaudal à esquerda

9.2.4 Tipos de Erro

Há dois tipos possíveis de erro ao testar uma hipótese estatística. Pode-se rejeitar uma hipótese quando ela é, de fato, verdadeira, ou aceitar uma hipótese quando ela é, de fato, falsa. A rejeição de uma hipótese verdadeira é chamada "erro tipo I". A aceitação de uma hipótese falsa constitui um "erro tipo II".

As probabilidades desses dois tipos de erros são designados, respectivamente, por α e β.

A probabilidade α do erro tipo I é denominada "nível de significância" do teste.

Os possíveis erros e acertos de um teste estão sintetizados na tabela:

		Realidade	
		H_o verdadeira	H_o falsa
Decisão	Aceitar H_0	Decisão correta $(1 - \alpha)$	Erro tipo II (β)
	Rejeitar H_0	Erro tipo I (α)	Decisão correta $(1 - \beta)$

Observe que o erro tipo I só poderá ser cometido se se rejeitar H_0, e o erro tipo II, quando se aceitar H_0.

O tomador de decisão deseja, obviamente, reduzir ao mínimo as probabilidades dos dois tipos de erros. Infelizmente, esta é uma tarefa difícil, porque, para uma amostra de determinado tamanho, a probabilidade de se incorrer em um erro tipo II aumenta à medida que diminui a probabilidade do erro tipo I. E vice-versa. A redução simultânea dos erros poderá ser alcançada pelo aumento do tamanho da amostra.

9.2.5 Configuração sobre o Mecanismo dos Erros

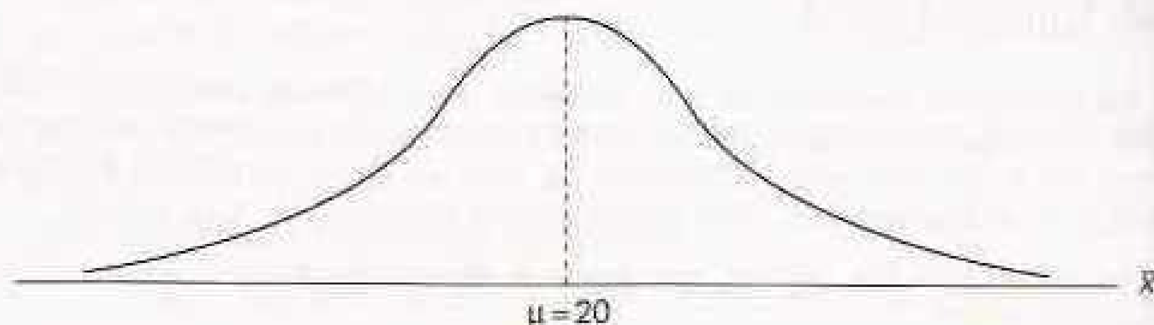
Para compreender o relacionamento dos erros e suas dimensões vamos idealizar um exemplo.

Deseja-se testar H_0: $\mu = 20$ contra H_1: $\mu > 20$. Sabe-se que a variância da população vale 16, e que foi retirada de uma amostra de 16 elementos. Ou seja:

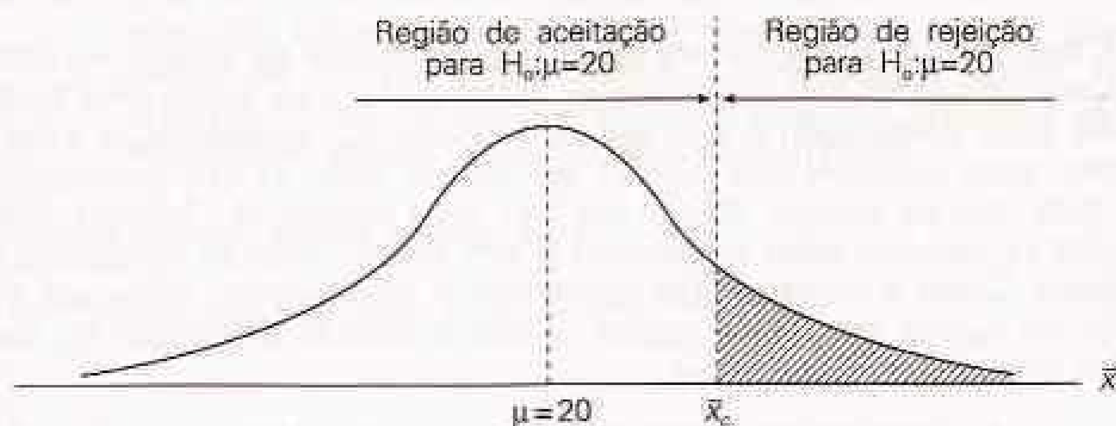
H_0: $\mu = 20$ $\qquad\qquad$ $\sigma^2 = 16$ $\qquad\qquad$ $n = 16$

H_1: $\mu > 20$

Como $\bar{x}$, estimador de μ, que por hipótese vale 20, tem-se graficamente:

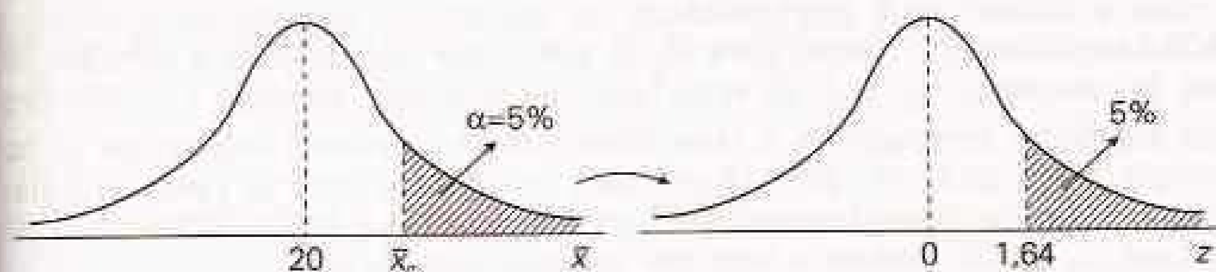


Ora, para valores amostrais de $\bar{x}$ próximos de 20 a hipótese H_0 poderá ser aceita. Como H_1: $\mu > 20$, deve-se ter um limite crítico à direita para valores de $\bar{x}$. Assim:



Observe: A área hachurada à direita de $\bar{x}_c$ corresponde à probabilidade de rejeitar H_0: $\mu = 20$, quando H_0: $\mu = 20$ é verdadeira. Ou seja, a área representa α (probabilidade de cometer o erro tipo I).

Para encontrar o limite crítico ($\bar{x}_c$) pode-se atribuir diversos valores para α. Admita que $\alpha = 5\%$. Assim, para se encontrar $\bar{x}_c$ basta passar para distribuição normal das médias para a distribuição normal padrão.



$$\bar{x} \overset{d}{=} N\left(\mu; \frac{\sigma^2}{n}\right) \rightarrow Z \overset{d}{=} N\ (0, 1)$$

$$Z = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{ou} \quad 1{,}64 = \frac{\bar{x}_c - 20}{\frac{4}{\sqrt{16}}}$$

logo: $\bar{x}_c = 21{,}64$.

Assim, a regra de decisão para H_0 será:

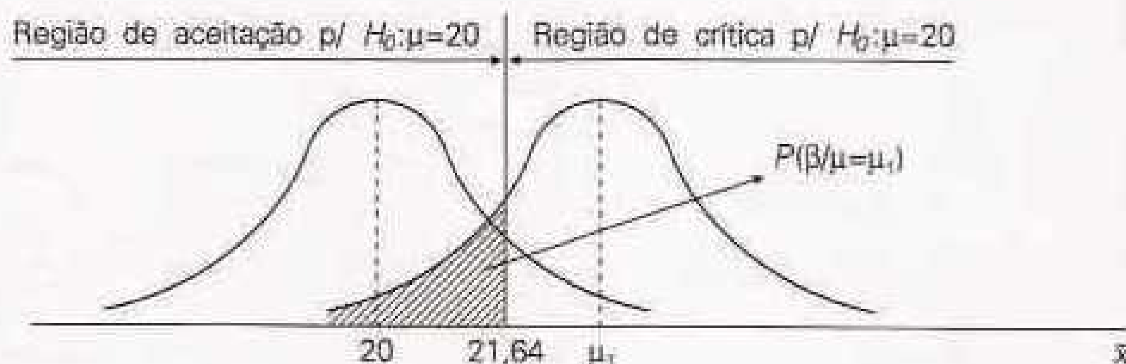
Rejeitar H_0 quando $\bar{x} > 21,64$

Aceitar H_0 quando $\bar{x} \leq 21,64$.

É fácil constatar que se tem grande probabilidade de aceitar H_0 (95%) e pouca probabilidade (5%) de rejeitar H_0. Ora, quando se aceita uma hipótese pode-se estar cometendo o erro tipo II – aceitar H_0, quando H_0 é falsa. No exemplo, essa probabilidade poderá ser de até 95%. O que preocupa. Por outro lado, tem-se apenas 5% de chances para rejeitar H_0. Todavia, quando se rejeita H_0 pode-se estar cometendo o erro tipo I – rejeitar H_0 quando H_0 é verdadeira. Como a probabilidade neste caso é relativamente baixa, até 5%, a decisão de rejeitar H_0 é mais segura do que a decisão de aceitar H_0. Esta é a lógica do teste de significância:

- atribuem-se baixos valores para α geralmente de 1% a 10%;
- formula-se H_0 com a pretensão de rejeitá-la, daí o nome de "hipótese nula";
- se o teste indicar a rejeição de H_0 tem-se um indicador mais seguro para a decisão;
- caso o teste indique a aceitação de H_0 diz-se que, com o nível de significância α, não se pode rejeitar H_0, e nestes casos a decisão não é tão segura quanto à rejeição de H_0.

Fixado α, pode-se determinar a probabilidade β de se cometer o erro tipo II. Para o cálculo de β (probabilidade de aceitar H_0, quando H_0 é falsa), é preciso admitir outros valores para H_0, já que o seu valor original é considerado falso. No exemplo, H_0: $\mu = 20$ seria falso, ou seja, em realidade $\mu > 20$. Ora, essa suposição corresponde a uma infinidade de possíveis valores de μ, por exemplo: 20,1; 20,5; 21; 22 ... Para cada um desses valores pode-se determinar o valor de β condicionado à hipótese admitida. Assim, para um valor qualquer, $\mu_1 > 20$, tem-se a seguinte configuração de β.

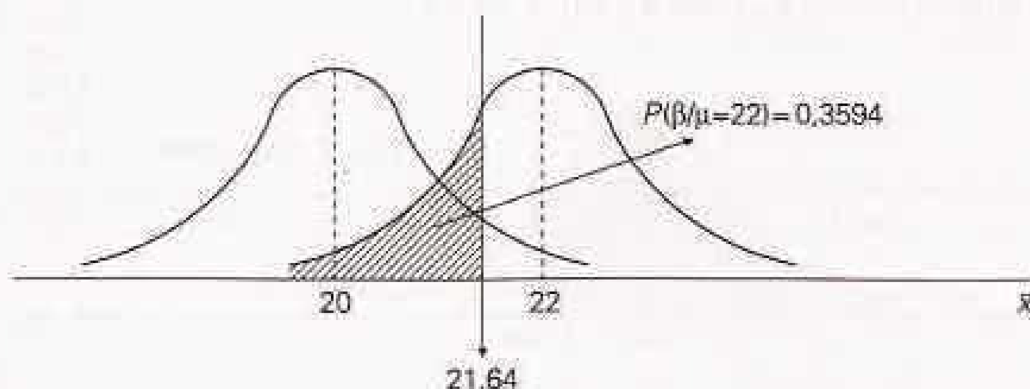


Se $\bar{x}$, por exemplo, for igual a 20,5, será aceita a hipótese falsa H_0: $\mu = 20$, quando na realidade a verdadeira hipótese é H_0: $\mu = \mu_1$.

Observe: Quando se consideram valores de μ_1 próximos de 20 tem-se elevados índices para β. Imagine, olhando o gráfico, o deslocamento de μ_1 para a esquerda. Quando $\mu_1 = 21{,}64$ tem-se $P(\beta/\mu_1) = 50\%$, e esse valor irá crescendo à medida que se consideram valores para μ_1 menores que 21,64.

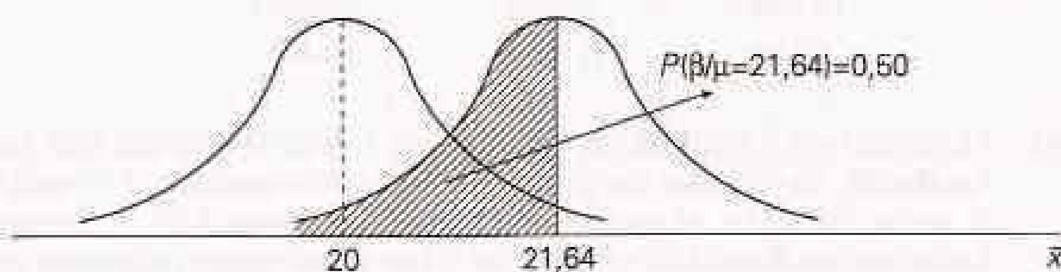
Vamos calcular alguns valores de β.

a) *Quanto valerá β, dado que $\mu = 22$?*



$$P(\beta/\mu = 22) = P(\bar{x} < 21{,}64/\mu = 22) = P\left(Z < \frac{21{,}64 - 22}{\frac{4}{\sqrt{16}}} = -0{,}36\right) = 0{,}3594$$

b) Quanto valerá β, dado que $\mu = 21{,}64$?



$$P(\beta/\mu = 21{,}64) = P(\bar{x} < 21{,}64/\mu = 21{,}64) = P\left(Z < \frac{21{,}64 - 21{,}64}{1} = 0\right) = 0{,}5000$$

Assim, a regra de decisão para H_0 será:

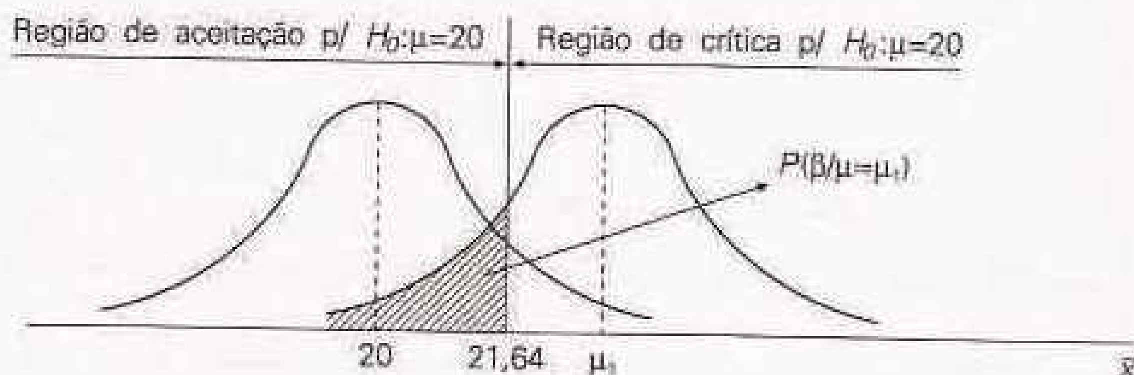
Rejeitar H_0 quando $\bar{x} > 21{,}64$

Aceitar H_0 quando $\bar{x} \leq 21{,}64$.

É fácil constatar que se tem grande probabilidade de aceitar H_0 (95%) e pouca probabilidade (5%) de rejeitar H_0. Ora, quando se aceita uma hipótese pode-se estar cometendo o erro tipo II – aceitar H_0, quando H_0 é falsa. No exemplo, essa probabilidade poderá ser de até 95%. O que preocupa. Por outro lado, tem-se apenas 5% de chances para rejeitar H_0. Todavia, quando se rejeita H_0 pode-se estar cometendo o erro tipo I – rejeitar H_0 quando H_0 é verdadeira. Como a probabilidade neste caso é relativamente baixa, até 5%, a decisão de rejeitar H_0 é mais segura do que a decisão de aceitar H_0. Esta é a lógica do teste de significância:

- atribuem-se baixos valores para α geralmente de 1% a 10%;
- formula-se H_0 com a pretensão de rejeitá-la, daí o nome de "hipótese nula";
- se o teste indicar a rejeição de H_0 tem-se um indicador mais seguro para a decisão;
- caso o teste indique a aceitação de H_0 diz-se que, com o nível de significância α, não se pode rejeitar H_0, e nestes casos a decisão não é tão segura quanto à rejeição de H_0.

Fixado α, pode-se determinar a probabilidade β de se cometer o erro tipo II. Para o cálculo de β (probabilidade de aceitar H_0, quando H_0 é falsa), é preciso admitir outros valores para H_0, já que o seu valor original é considerado falso. No exemplo, H_0: $\mu = 20$ seria falso, ou seja, em realidade $\mu > 20$. Ora, essa suposição corresponde a uma infinidade de possíveis valores de μ, por exemplo: 20,1; 20,5; 21; 22 ... Para cada um desses valores pode-se determinar o valor de β condicionado à hipótese admitida. Assim, para um valor qualquer, $\mu_1 > 20$, tem-se a seguinte configuração de β.

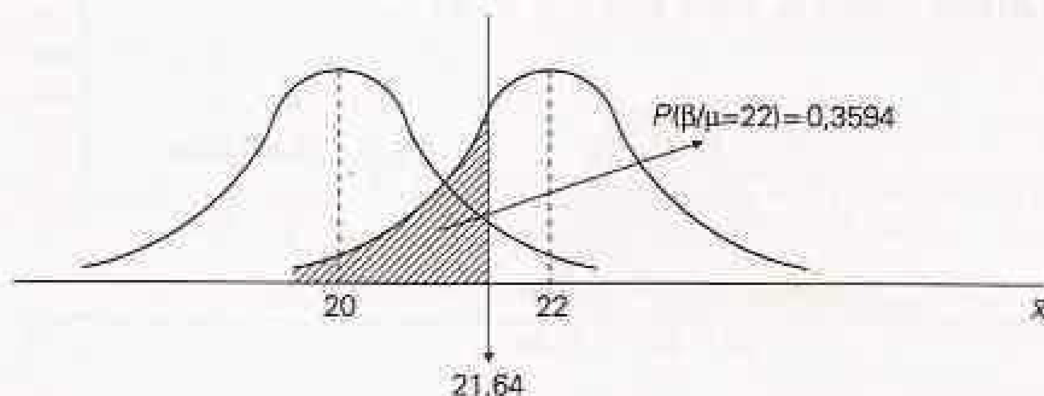


Se $\bar{x}$, por exemplo, for igual a 20,5, será aceita a hipótese falsa H_0: $\mu = 20$, quando na realidade a verdadeira hipótese é H_0: $\mu = \mu_1$.

Observe: Quando se consideram valores de μ_1 próximos de 20 tem-se elevados índices para β. Imagine, olhando o gráfico, o deslocamento de μ_1 para a esquerda. Quando $\mu_1 = 21{,}64$ tem-se $P(\beta/\mu_1) = 50\%$, e esse valor irá crescendo à medida que se consideram valores para μ_1 menores que 21,64.

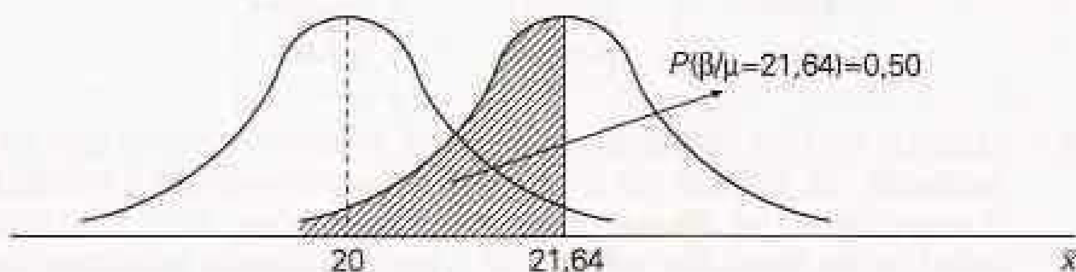
Vamos calcular alguns valores de β.

a) Quanto valerá β, dado que $\mu = 22$?



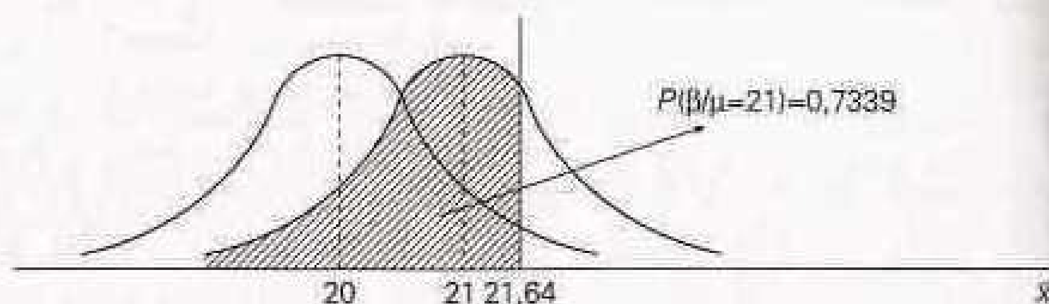
$$P(\beta/\mu = 22) = P(\bar{x} < 21{,}64/\mu = 22) = P\left(Z < \frac{21{,}64 - 22}{\frac{4}{\sqrt{16}}} = -0{,}36\right) = 0{,}3594$$

b) Quanto valerá β, dado que $\mu = 21{,}64$?



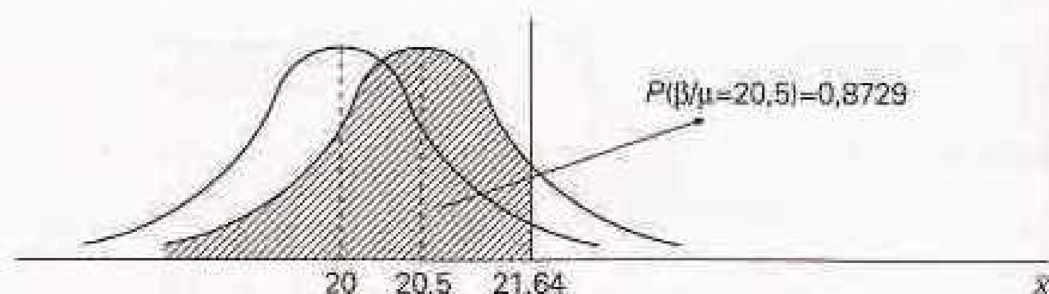
$$P(\beta/\mu = 21{,}64) = P(\bar{x} < 21{,}64/\mu = 21{,}64) = P\left(Z < \frac{21{,}64 - 21{,}64}{1} = 0\right) = 0{,}5000$$

c) Quanto valerá β, dado que μ = 21?



$$P(\beta/\mu = 21) = P(\bar{x} < 21{,}64/\mu = 21) = P\left(Z < \frac{21{,}64 - 21}{1} = 0{,}64\right) = 0{,}7339$$

d) Quanto valerá β, dado que μ = 20,5?



$$P(\beta/\mu = 20{,}5) = P(\bar{x} < 21{,}64/\mu = 21) = P\left(Z < \frac{21{,}64 - 20{,}5}{1} = 1{,}14\right) = 0{,}8729$$

Assim:

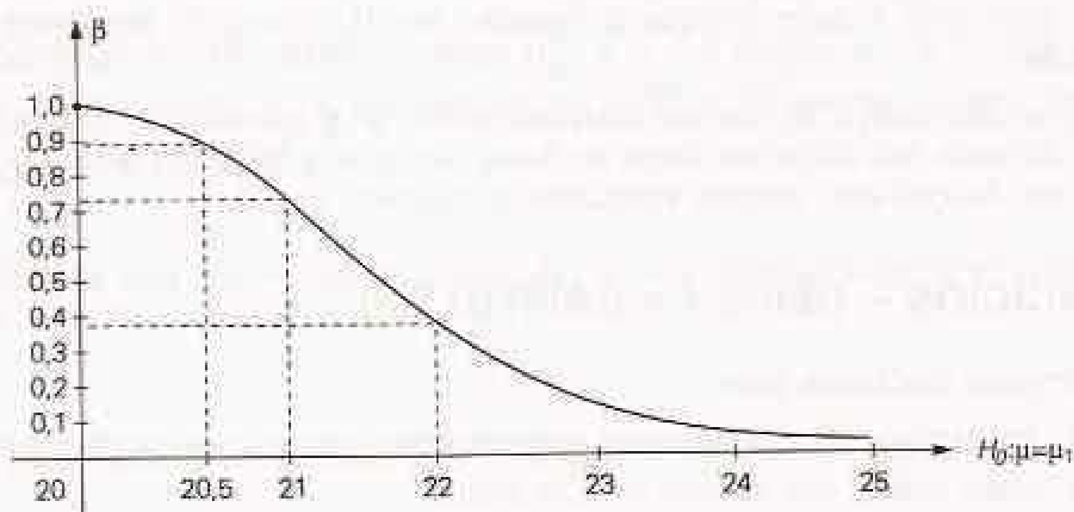
H_0: $\mu = \mu_1$	β	1 − β
20,50	0,8729	0,1271
21,00	0,7389	0,2611
21,64	0,5000	0,5000
22,00	0,3594	0,6406

Observe: Quando se tem hipóteses próximas à hipótese original que se está testando, os valores de β são elevados, diminuindo à medida que o valor de μ_1 se afasta do valor testado. Veja. Não é comprometedor se ter elevados valores de β para hipóteses próximas daquelas que se está testando. Ou seja, dizer que há um erro de 0,8729 (87,29%) de se aceitar H_0 : μ = 20, quando na realidade H_0: μ = 20,5 não é um resultado preocupante, pois se está aceitando 20, quando o verdadeiro poderá ser 20,5. A diferença de 0,5 (erro absoluto) não é tão preocupante. Concordas?

9.2.6 Curva Característica de Operação (CCO)

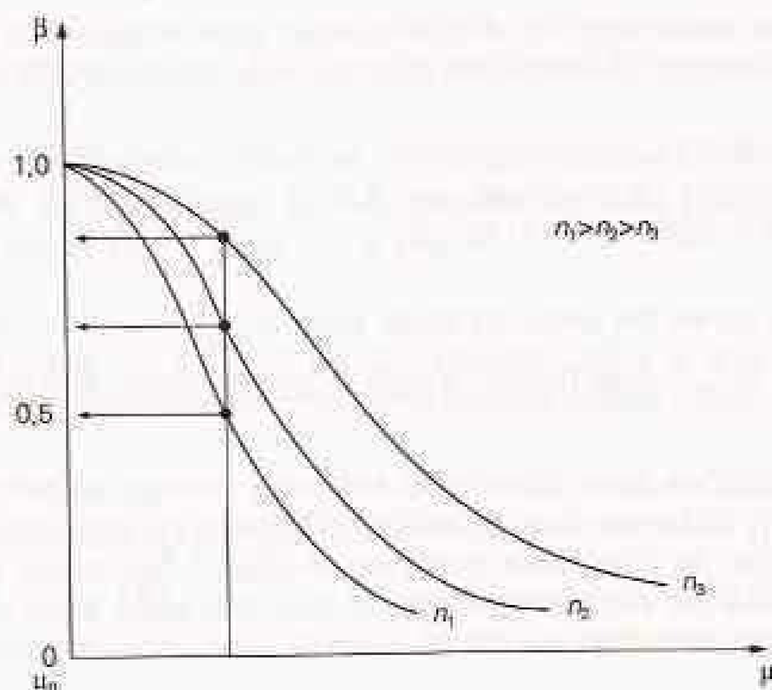
Trata-se da construção de uma curva que expressa o comportamento do erro β em função das diversas hipóteses alternativas feitas para H_0, fixando-se o nível de α.

No exemplo que se está analisando, a CCO é dado por:



CCO – H_0:μ=20 – α=5%; σ^2=16 e n=16

Como era de esperar, quando se aumenta o tamanho da amostra conseguem-se menores erros para β, admitindo-se α baixo (entre 1% e 10%). O gráfico a seguir exibe tal situação.



Como você já percebeu, as curvas características de operação constituem elementos para análise do comportamento do erro β, e também α, pois é possível construir curvas para combinações α e β, variando-se os tamanhos amostrais.

Para tomar decisões inequívocas, ou seja, sabendo-se previamente quais os erros β e α, é preciso a construção das CCO. Todavia, como é moroso o trabalho dos cálculos dos erros β, na prática, realizam-se mais freqüentemente os "testes de significância". Ou seja, considera-se apenas o risco α, "torcendo-se" para que o teste indique a rejeição de H_0, como foi explicado anteriormente.

Uma alternativa às curvas características é a construção da *curva de poder do teste*, ou *curva de força do teste*, em que a força do teste $(1 - \beta)$ é obtida em função dos valores atribuídos à hipótese H_0.

EXERCÍCIOS – SÉRIE I – CAPÍTULO 9

1. Formule hipóteses para:
   a) porcentagem de brasileiros analfabetos;
   b) peso médio dos alunos de sua classe;
   c) variância dos salários dos professores da sua escola;
   d) salário médio dos trabalhadores brasileiros;
   e) proporção de garotos da sua escola.

2. Explique os erros tipo I e II que podem ocorrer quando um professor decide aprovar ou reprovar um aluno.

3. Explique os erros tipo I e II que podem ocorrer quando um gerente de recursos humanos decide contratar, ou não, determinado profissional.

4. Construa a CCO para se testar H_0: $\mu = 50$, contra H_1: $\mu > 50$, admitindo-se intervalos sucessivos de 0,4 a partir de 50 até 55, sendo $\sigma^2 = 25$; $n = 25$ e $\alpha = 10\%$.

5. Construa a curva de força do teste para H_0: $\mu = 15$, H_1: $\mu \neq 15$, em que $n = 9$; $\sigma = 2$ e $\alpha = 5\%$. Admita: 12,0; 12,5; 13,0; 13,5; 14,0; 14,5; 15,0; 15,5; 16,0; 16,5; 17,0; 17,5; 18,0 como possíveis hipóteses verdadeiras.

6. As especificações para fabricação de certo produto exigem 25 mg de um composto *Y*. Sabe-se que a distribuição de *Y* no produto é normal, com desvio-padrão de 2 mg para qualquer média. O fabricante está examinando a inspeção de cada lote antes de sua liberação para se decidir entre entregá-lo ao consumo ou mantê-lo para reprocessamento.

a) Para uma amostra de 16, quais seriam os limites decisórios se a probabilidade de reter incorretamente o produto que satisfaça as especificações deva ser limitado a 5%?

b) Trace a curva de força para a norma decisória em intervalos de 0,40 mg de $\mu = 23,0$ até $\mu = 27,0$.

c) Trace a curva de força para amostras de tamanho 16 com $\alpha = 20\%$.

d) Trace a curva de força para amostras de 36 com $\alpha = 5\%$.

7. Construir o CCO para se testar H_0: $p = 0,5$ contra H_1: $p > 0,5$, sendo $n = 100$. Admita $p = 0,55; 0,60; 0,65; 0,70; 0,75$, sendo $\alpha = 5\%$. Lembre-se: $f \stackrel{d}{=} N\left(p, \frac{pq}{n}\right)$.

8. Quanto vale $\beta/\mu = 1,65$ m, sendo H_0: $\mu = 1,70$ m, $n = 36$; e $\sigma^2 = 0,04$ m^2 e $\alpha = 8\%$. H_1: $\mu < 1,70$.

9.3 TESTES DE SIGNIFICÂNCIA

Como já foi dito, os testes de significância consideram apenas o erro α. São os mais usados nas pesquisas educacionais, sócio-econômicas. . .

O procedimento para realização dos testes de significância é resumido nos seguintes passos:

1. enunciar as hipóteses H_0 e H_1;
2. fixar o limite do erro α, e identificar a variável do teste;
3. com o auxílio das tabelas estatísticas, considerando α e a variável do teste, determinar as RC (região crítica) e RA (região de aceitação) para H_0;
4. com os elementos amostrais, calcular o valor da variável do teste;
5. concluir pela aceitação ou rejeição de H_0 pela comparação do valor obtido no 4º passo com RA e RC.

9.3.1 Teste de Significância para Médias

1. H_0: $\mu = \mu_0$

   H_1: uma das alternativas

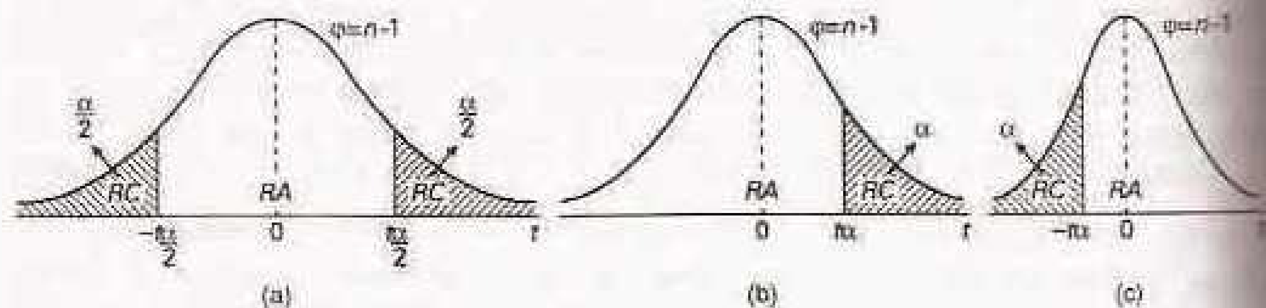
   $\mu \neq \mu_0$ (a)

   $\mu > \mu_0$ (b)

   $\mu < \mu_0$ (c)

2. Fixar α. Admitindo-se que σ^2 é desconhecida, a variável do teste será "t" de Student, como $\varphi = (n - 1)$.

3. Com auxílio da tabela "t" determinam-se RA e RC.



4. Cálculo do valor da variável

$$t_{cal} = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$$

onde:
$\bar{x}$ = média amostral
μ_0 = valor da hipótese nula
S = desvio-padrão amostral
n = tamanho da amostra

5. Conclusões

a) Se $-t\frac{\alpha}{2} \leq t_{cal} \leq t\frac{\alpha}{2}$ não se pode rejeitar H_0

Se $t_{cal} > t\frac{\alpha}{2}$ ou $t_{cal} < -t\frac{\alpha}{2}$, rejeita-se H_0.

b) Se $t_{cal} < t\alpha$, não se pode rejeitar H_0

Se $t_{cal} > t\alpha$, rejeita-se H_0

c) Se $t_{cal} > t\alpha$, não se pode rejeitar H_0

Se $t_{cal} < t\alpha$, rejeita-se H_0

Exemplo: Os dois registros dos últimos aos de um colégio, atestam para os calouros admitidos uma nota média 115 (teste vocacional). Para testar a hipótese de que a média de uma nova turma é a mesma, tirou-se, ao acaso, uma amostra de 20 notas, obtendo-se média 118 e desvio-padrão 20. Admitir que $\alpha = 0,05$, para efetuar o teste. Então,

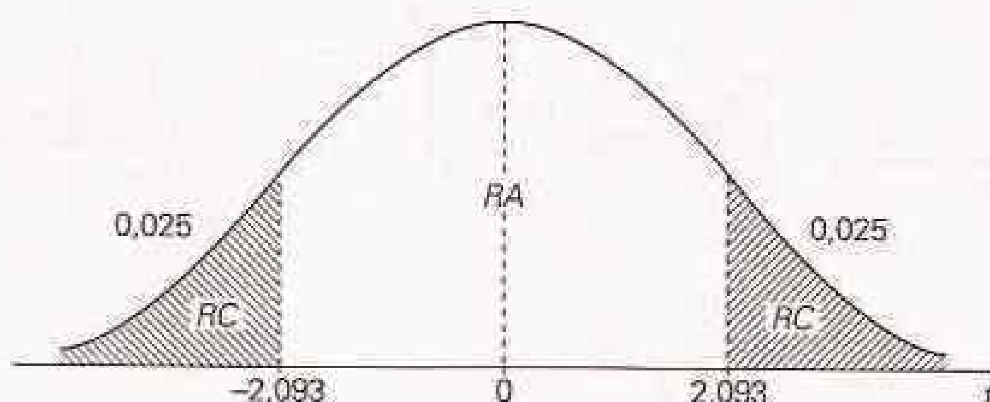
1º) H_0: $\mu = 115$

H_1: $\mu \neq 115$

2º) $\alpha = 0{,}05$

Variável "t" com 19 graus de liberdade.

3º)



4º) $t_{cal} = \dfrac{118 - 115}{\dfrac{20}{\sqrt{20}}} = 0{,}67$

5º) Como $-2{,}0930 \leq t_{cal} \leq 2{,}0930$, não se pode rejeitar H_0: $\mu = 115$ com esse nível de significância.

9.3.2 Teste de Significância para Variâncias

1. H_0: $\sigma^2 = \sigma_0^2$

   H_1: uma das alternativas

   $\sigma^2 \neq \sigma_0^2$ (a)

   $\sigma^2 > \sigma_0^2$ (b)

   $\sigma^2 < \sigma_0^2$ (c)

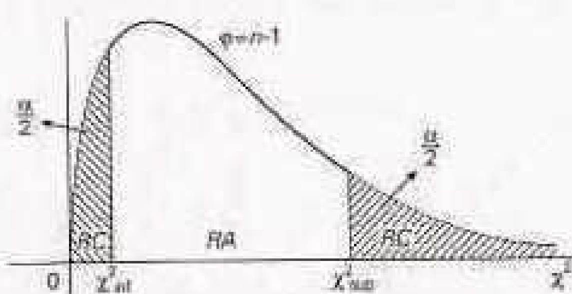
2. Fixar α. Escolher a variável qui-quadrado com $\varphi = (n - 1)$.

3. Com auxílio da tabela "χ^2" determinam-se RA e RC. Vide gráficos *a*, *b*, e *c* (página seguinte).

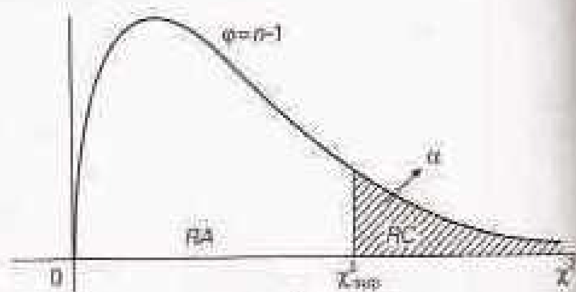
4. Cálculo do valor da variável

$$\chi^2_{cal} = \frac{(n-1)S^2}{\sigma_0^2}$$

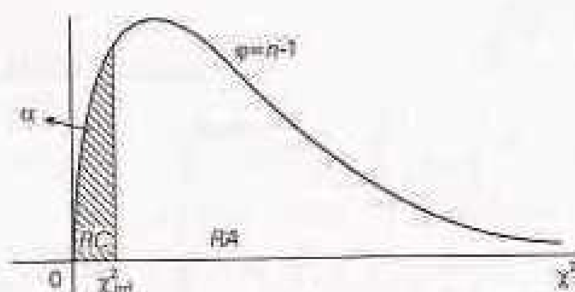
onde:
n = tamanho da amostra
S^2 = variância amostral
σ^2 = valor da hipótese nula



(a)



(b)



(c)

5. Conclusões:

   a) Se $\chi^2_{inf} \leq \chi^2_{cal} \leq \chi^2_{sup}$, não se pode rejeitar H_0.

      Se $\chi^2_{cal} > \chi^2_{sup}$ ou $\chi^2_{cal} < \chi^2_{inf}$, rejeita-se H_0.

   b) Se $\chi^2_{cal} < \chi^2_{sup}$, não se pode rejeitar H_0.

      Se $\chi^2_{cal} > \chi^2_{sup}$, rejeita-se H_0.

   c) Se $\chi^2_{cal} > \chi^2_{inf}$, não se pode rejeitar H_0.

      Se $\chi^2_{cal} < \chi^2_{inf}$, rejeita-se H_0.

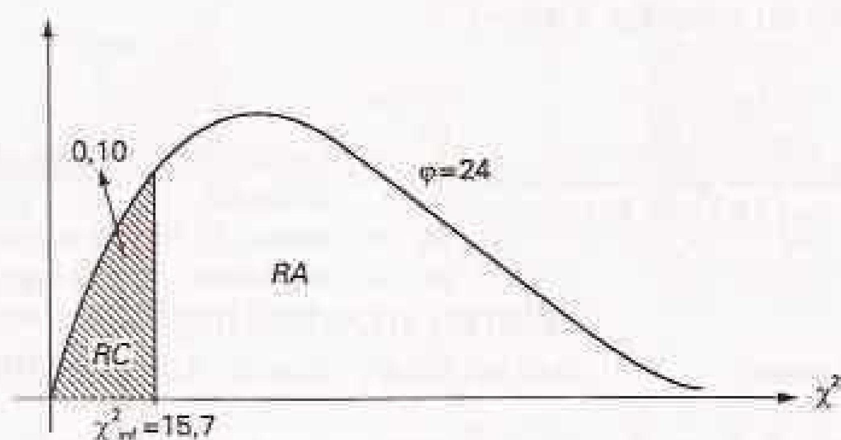
Exemplo: Para testar a hipótese de que a variância de uma população é 25, tirou-se uma amostra aleatória de 25 elementos obtendo-se $S^2 = 18,3$. Admitindo-se $\alpha = 0,10$, efetuar o teste de significância unicaudal à esquerda.

1º) H_0: $\sigma^2 = 25$

H_1: $\sigma^2 < 25$

2º) $\alpha = 0{,}10$; variável χ^2 com $25 - 1 = 24$ gℓ.

3º)



4º) $\chi^2_{cal} = \dfrac{(25 - 1) \cdot 18{,}3}{25} = 17{,}56$

5º) Como $\chi^2_{cal} > 15{,}7$, não se pode rejeitar $H_0 : \sigma^2 = 25$ ao nível de significância de 10%.

9.3.3 Teste de Significância para Proporções

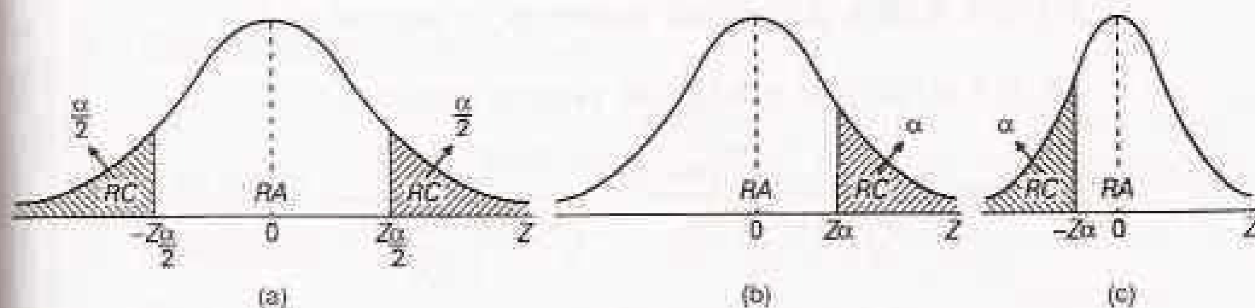
1. H_0: $p = p_0$

   H_1: uma das alternativas

   $p \neq p_0$ (a)

   $p > p_0$ (b)

   $p < p_0$ (c)



2. Fixar α. Escolher a variável norma padrão: Z.

3. Com auxílio da tabela da distribuição normal padrão, determinam-se RA e RC.

4. Cálculo do valor da variável

$$Z_{cal} = \frac{f - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

onde:

f = freqüência relativa do evento na amostra

p_0 = valor da hipótese nula

n = tamanho da amostra

5. Conclusões:

a) Se $-Z\frac{\alpha}{2} \leq Z_{cal} \leq Z\frac{\alpha}{2}$, não se pode rejeitar H_0

Se $Z_{cal} > Z\frac{\alpha}{2}$ ou $Z_{cal} < -Z\frac{\alpha}{2}$, rejeita-se H_0

b) Se $Z_{cal} < Z\alpha$, não se pode rejeitar H_0

Se $Z_{cal} > Z\alpha$, rejeita-se H_0

c) Se $Z_{cal} > -Z\alpha$, não se pode rejeitar H_0

Se $Z_{cal} < -Z\alpha$, rejeita-se H_0

Exemplo: As condições de mortalidade de uma região são tais que a proporção de nascidos que sobrevivem até 60 anos é de 0,6. Testar essa hipótese ao nível de 5% se em 1000 nascimentos amostrados aleatoriamente, verificou-se 530 sobreviventes até 60 anos.

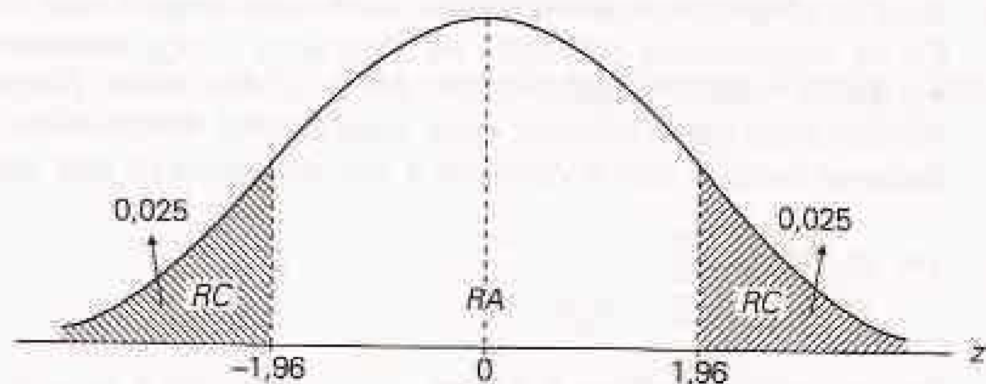
1º) H_0: $p = 0{,}6$
H_1: $p \neq 0{,}6$

2º) $\alpha = 0{,}05$ e a variável escolhida, a normal (0,1).

3º) RA e RC (ver gráfico na página seguinte).

4º) $$Z_{cal} = \frac{f - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0{,}53 - 0{,}6}{\sqrt{\frac{0{,}6(1 - 0{,}6)}{1000}}} = -4{,}42$$

5º) Como $Z_{cal} < -1{,}96$, rejeita-se H_0, concluindo-se, ao nível de 5%, que $p \neq 0{,}6$.



9.3.4 Teste de Significância para a Igualdade de duas Variâncias

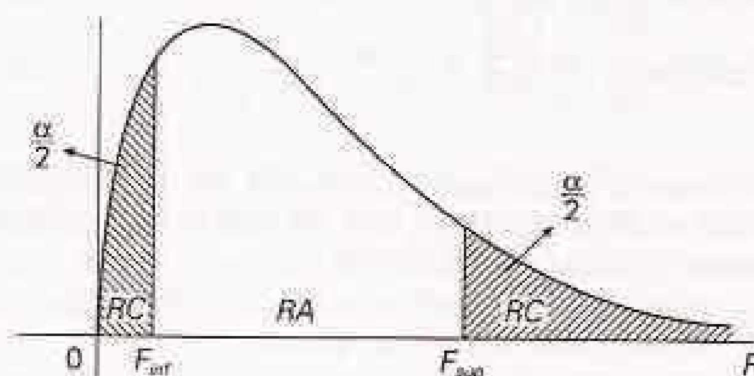
1. H_0: $\sigma_1^2 = \sigma_2^2$

   H_1: $\sigma_1^2 \neq \sigma_2^2$

   Podem-se testar as desigualdades < ou >, todavia, o mais comum é $\neq$.

2. Fixar α. Escolher a variável "F" com $(n_1 - 1)$ graus de liberdade no numerador, e $(n_2 - 1)$ graus de liberdade no denominador.

3. Com auxílio da tabela da distribuição "F", determinam-se RA e RC



4. Cálculo do valor da variável

$$F_{cal} = \frac{S_1^2}{S_2^2}$$

5. Conclusões:

   Se $F_{inf} \leq F_{cal} \leq F_{sup}$, não se pode rejeitar H_0.

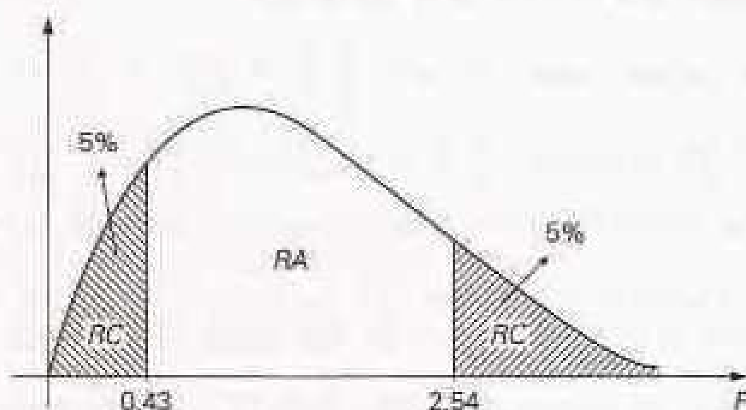
   Se $F_{cal} > F_{sup}$, ou $F_{cal} < F_{inf}$, rejeita-se H_0.

Exemplo: Dois programas de treinamento de funcionários foram efetuados. Os 21 funcionários treinados no programa antigo apresentam uma variância 146 em suas taxas de erro. No novo programa, 13 funcionários apresentaram uma variância de 200. Sendo $\alpha = 0{,}10$ pode-se concluir que a variância é diferente para os dois programas?

1º) H_0: $\sigma_1^2 = \sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$

2º) $\alpha = 0{,}10$. A variável é F com $\varphi_1 = n_1 - 1 = 20$ e $\varphi_2 = n_2 - 1 = 12$

3º) RA e RC



4º) $F_{cal} = \dfrac{S_1^2}{S_2^2} = \dfrac{146}{200} = 0{,}73$

5º) Como $0{,}43 \leq F_{cal} \leq 2{,}54$, não se pode rejeitar H_0, portanto, não se pode concluir que as variâncias sejam diferentes com esse nível de significância.

9.3.5 Teste de Significância para a Igualdade de duas Médias

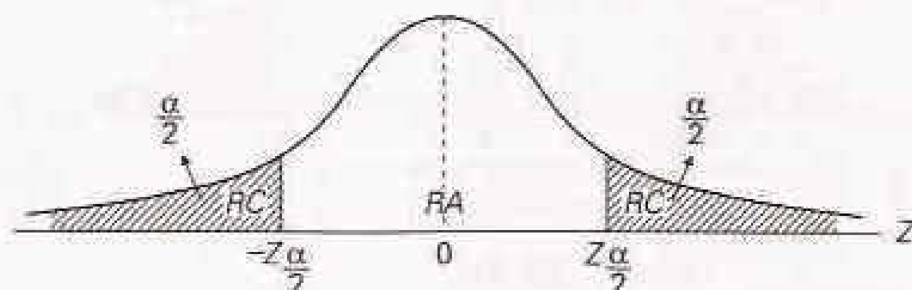
1º Caso: As variâncias populacionais são conhecidas, independentes e normais.

1. H_0: $\mu_1 = \mu_2$ ou $\mu_1 - \mu_2 = d$ onde $d > 0$ é uma diferença admitida entre as médias.

   H_1: $\mu_1 \neq \mu_2$ ou $\mu_1 - \mu_2 \neq d$
   Os testes unicaudais são permitidos.

2. Fixar α. Escolher a variável normal padrão: Z.

3. Com auxílio da tabela da distribuição normal padrão, determinar RA e RC.



4. Cálculo do valor da variável

$$Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5. Conclusões:

Se $-Z\frac{\alpha}{2} \leq Z_{cal} \leq Z\frac{\alpha}{2}$ não se pode rejeitar H_0.

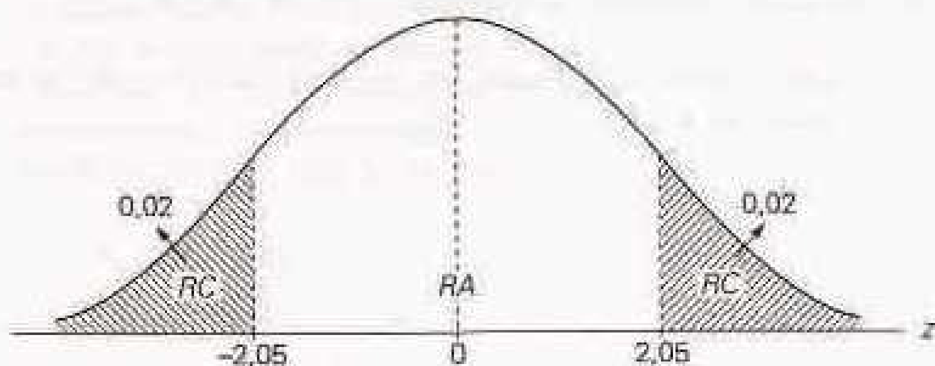
Se $Z_{cal} > Z\frac{\alpha}{2}$ ou $Z_{cal} < -Z\frac{\alpha}{2}$, rejeita-se H_0.

Exemplo: Um fabricante de pneus faz dois tipos. Para o tipo A, $\sigma = 2500$ milhas, e para o tipo B, $\sigma = 3000$ milhas. Um táxi testou 50 pneus do tipo A e 40 de tipo B, obtendo 24000 milhas e 26000 milhas de duração média dos respectivos tipos. Adotando-se um risco alfa de 4%, testar a hipótese de que a vida média dos dois tipos é a mesma.

1º) H_0: $\mu_A - \mu_B = 0 \therefore \mu_A = \mu_B$
H_1: $\mu_A \neq \mu_B$

2º) $\alpha = 0,04$, a variável é $N(0,1)$

3º) RA e RC



4º) $z = \dfrac{24000 - 26000}{\sqrt{\dfrac{(2500)^2}{50} + \dfrac{(3000)^2}{40}}} = -3{,}38$

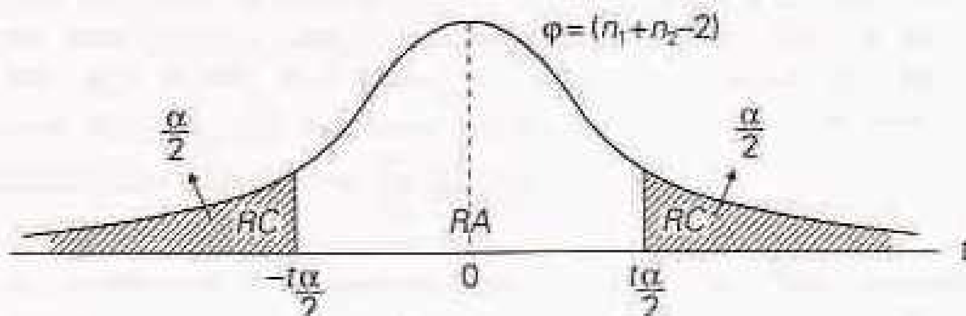
5º) Como $Z_{cal} < -2{,}05$, rejeita-se H_0, concluindo-se com risco de 4% que as vidas médias dos pneus são diferentes.

2º caso: As variâncias populacionais são desconhecidas e, admitidas iguais, independentes e normais.

1. H_0: $\mu_1 = \mu_2$ ou $\mu_1 - \mu_2 = d$
   H_1: $\mu_1 \neq \mu_2$ ou $\mu_1 - \mu_2 \neq d$

2. Fixar α. Escolher a variável "t" com $\varphi = (n_1 + n_2 - 2)$

3. Com auxílio da tabela da distribuição "t", determinam-se RA e RC.



4. Cálculo do valor da variável

$$t_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - d}{S_c \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}}$$

onde, S_c = desvio-padrão comum é dado por:

$$S_c = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

5. Conclusões:

Se $-t\frac{\alpha}{2} \leq t_{cal} \leq t\frac{\alpha}{2}$, não se pode rejeitar H_0.

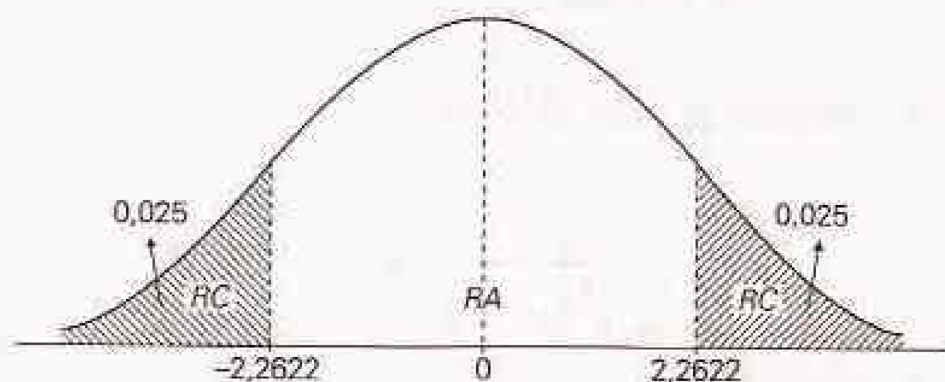
Se $t_{cal} > t\frac{\alpha}{2}$ ou $t_{cal} < -t\frac{\alpha}{2}$, rejeita-se H_0.

Exemplo: Dois tipos de tinta foram testados sob as mesmas condições meteorológicas. O tipo *A* registrou uma média de 80 com um desvio de 5 em 5 partes. O tipo *B*, uma média de 83 com um desvio de 4 em 6 partes. Adotando $\alpha = 0{,}05$ testar a hipótese da igualdade das médias.

1º) $H_0: \mu_A = \mu_B$
$H_1: \mu_A \neq \mu_B$

2º) $\alpha = 0{,}05$ e a variável t com $5 + 6 - 2 = 9$ graus de liberdade.

3º) RA e RC



4º) $S_c = \sqrt{\dfrac{(5-1)5^2 + (6-1)4^2}{5+6-2}} = 4{,}47$

$$t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S_c \sqrt{\dfrac{n_1 + n_2}{n_1 \cdot n_2}}} = \frac{80 - 83}{4{,}47 \sqrt{\dfrac{5+6}{30}}} = -1{,}12$$

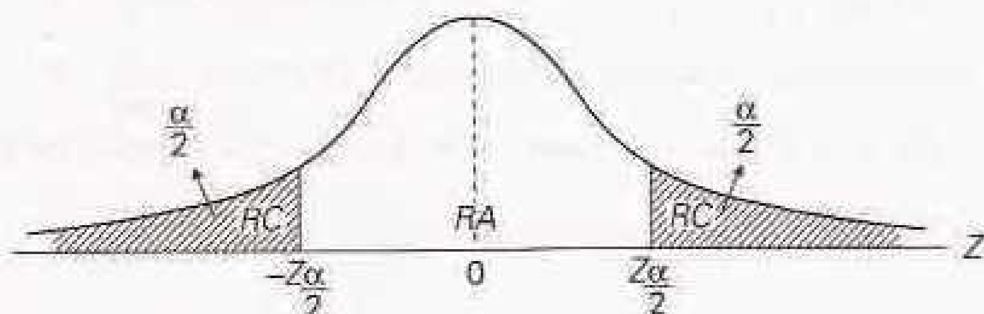
5º) Como $-2{,}2622 \leq t_{cal} \leq 2{,}2622$, não se pode rejeitar H_0 com esse nível de significância.

9.3.6 Teste de Significância para a Igualdade de Duas Proporções

1. H_0: $p_1 = p_2$
   H_1: $p_1 \neq p_2$

2. Fixar α. Escolher a variável normal padrão: Z.

3. Com auxílio da tabela da distribuição normal padrão, determinam-se RA e RC.



4. Cálculo do valor da variável

$$Z_{cal} = \frac{f_1 - f_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

onde: f_1 e f_2 são as freqüências relativas amostrais; $\hat{p}$ é o estimado comum a p_1 e p_2, dado por:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$f_1 = \frac{x_1}{n_1} \quad \text{e} \quad f_2 = \frac{x_2}{n_2}$$

5. Conclusões:

Se $-Z\frac{\alpha}{2} \leq Z_{cal} \leq Z\frac{\alpha}{2}$, não se pode rejeitar H_0.

Se $Z_{cal} > Z\frac{\alpha}{2}$ ou $Z_{cal} < -Z\frac{\alpha}{2}$, rejeita-se o H_0.

Exemplo: Deseja-se testar se são iguais as proporções de homens e mulheres que lêem revista e se lembram de determinado anúncio. São os seguintes os resultados de amostras aleatórias independentes de homens e mulheres:

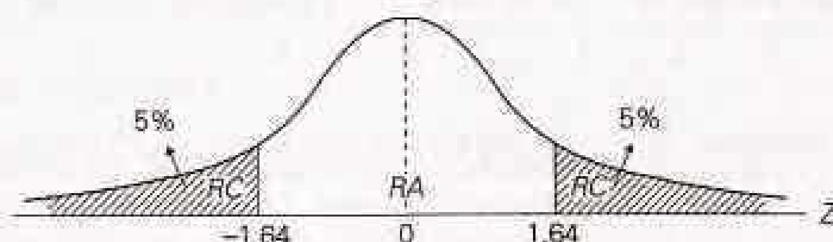
Homens	***Mulheres***
$x_1 = 70$	$x_2 = 50$
$n_1 = 200$	$n_2 = 200$

onde x_1 é o número de homens que se lembram do anúncio e x_2 é o correspondente número de mulheres. Admita $\alpha = 10\%$.

1. H_0: $p_1 = p_2$
   H_1: $p_1 \neq p_2$

2. $\alpha = 10\%$. Escolher a variável $N(0,1)$.

3. Determinação de RA e RC.



4. $f_1 = \dfrac{x_1}{n_1} = \dfrac{70}{200} = 0{,}35$ $\qquad f_2 = \dfrac{x_2}{n_2} = \dfrac{50}{200} = 0{,}25$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{70 + 50}{200 + 200} = 0{,}30$$

$$Z_{\text{cal}} = \frac{0{,}35 - 0{,}25}{\sqrt{0{,}3\,(0{,}7)\left(\dfrac{2}{200}\right)}} = 2{,}18$$

5. Conclusão:

   Como $Z_{\text{cal}} > 1{,}64$, rejeita-se a hipótese da igualdade das proporções, concluindo-se com risco de 10% que as proporções são diferentes.

EXERCÍCIOS – SÉRIE II – CAPÍTULO 9

Testes para a Média Populacional

1. Uma amostra de 25 elementos resultou média 13,5 com desvio-padrão 4,4. Efetuar o teste ao nível de 0,05 para a hipótese que $\mu = 16$ contra $\mu \neq 16$; e $\mu < 16$.

2. Retirada uma amostra de 15 parafusos, obteve-se as seguintes medidas para seus diâmetros:

   10 10 10 11 11 12 12 12 12
   13 13 14 14 14 15

   Testar H_0: $\mu = 12{,}5$ contra $\mu \neq 12{,}5$; $\mu > 12{,}5$; $\mu < 12{,}5$. Adotando $\alpha = 0{,}05$.

3. Dada a distribuição amostral

Classes	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
F_i	3	5	8	3	2

   Testar a hipótese de que a média da população é 20 contra H_1: $\mu \neq 20$, sendo $\alpha = 2{,}5\%$.

4. As estaturas de 20 recém-nascidos foram tomadas no Departamento de Pediatria da FMRP, cujos resultados são em cm:

41	50	52	49	49	54	50	47	52	49
50	52	50	47	49	51	46	50	49	50

a) Suponha inicialmente que a população das estaturas é normal com variância 2 cm^2; teste a hipótese de que a média desta normal é 50 cm ($\alpha = 0{,}05$) (teste unicaudal).

b) Faça o mesmo teste para a média, mas agora desconhecendo a variância (teste unicaudal).

5. 15 animais foram alimentados com uma certa dieta durante 3 semanas e verificou-se os seguintes aumentos de pesos:

25	30	32	24	40	34	37	33
34	28	30	32	38	29	31	

Testar a hipótese de que a média é 30, sendo $\alpha = 10\%$ (teste bicaudal).

Testes para a Variância Populacional

6. Um laboratório fez oito determinações da quantidade de impurezas em porções de certo composto. Os valores eram: 12,4; 12,6; 12,0; 12,0; 12,1; 12,3; 12,5 e 12,7 mg.

a) Estimar a variância de impurezas entre porções.

b) Testar a hipótese de que a variância é 1, ao nível de $\alpha = 0{,}05$ e $\alpha = 0{,}10$, contra H_1: $\sigma^2 < 1$.

7. Testar H_0: $\sigma^2 = 10$ contra H_1: $\sigma^2 \neq 10$ admitindo a seguinte distribuição:

Classes	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
F_i	3	5	8	3	1

Admita $\alpha = 20\%$

8. Suponha $X = N(\mu, \sigma^2)$ em que μ e σ^2 são desconhecidos. Uma amostra de tamanho 15 forneceu $\Sigma x_i = 8{,}7$ e $\Sigma x_i^2 = 27{,}3$. Testar a hipótese de que a variância da população é 4. Adotar $\alpha = 1\%$ (testes uni e bicaudal).

Testes para a Proporção

9. Uma amostra de 500 eleitores selecionados ao acaso dá 52% ao Partido Democrático. Poderia esta amostra ter sido retirada de uma população que tivesse 50% de eleitores democratas? Admita $\alpha = 0{,}05$.

10. Com base na tabela

	Cigarros s/filtro	*Cigarros c/filtro*	*Não fumam*	*Total*
Homens	12	64	14	90
Mulheres	8	26	16	50
Total	20	90	30	140

a) Testar a hipótese de que a proporção de fumantes é 80% sendo α = 0,0[illegible]

b) Testar a hipótese de que a proporção dos que fumam cigarros com filtro é 70%. $\alpha = 0{,}02$.

c) Testar a hipótese de que a população de fumantes femininas é 40% $\alpha = 0{,}01$.

11. Lança-se uma moeda 100 vezes e obtém-se 60 caras. Testar ao nível de 5% a hipótese de que a moeda é honesta.

12. Uma pesquisa revelou que das 500 donas de casas consultadas, 30[illegible] preferiram o detergente *A*. Testar a hipótese ao nível de 0,04 para H_0: $p = 0{,}5$ contra H_1: $p \neq 0{,}5$.

13. A experiência tem demonstrado que 40% dos estudantes são reprovados num exame de ESTATÍSTICA (?). Se 40 de 90 estudantes fossem reprovados, poderíamos concluir que esses estudantes são inferiores em Estatística? $\alpha = 5\%$.

Testes para a Igualdade de Duas Variâncias

14. Para verificar a eficácia de uma nova droga foram injetadas doses em 72 ratos, obtendo-se a seguinte tabela:

	tamanho da amostra	*Variância*
Machos	41	43,2
Fêmeas	31	29,5

Testar a igualdade das variâncias sendo $\alpha = 10\%$.

15. Tome duas dietas quaisquer do exercício 19 e teste a igualdade das variâncias sendo $\alpha = 10\%$.

16. Sendo $\alpha = 10\%$, testar a igualdade das variâncias para as duas marcas *A* e *B* do exercício 18.

Testes para a Igualdade entre Duas Médias

17. Sendo: Amostra 1: $n_1 = 60$; $\bar{x} = 5{,}71$; $\sigma^2 = 43$
Amostra 2: $n_2 = 35$; $\bar{x} = 4{,}12$; $\sigma^2 = 28$

testar H_0: $\mu_1 = \mu_2$ sendo $\alpha = 0{,}04$.

18. Na tabela abaixo estão registrados os índices de vendas em 6 supermercados para os produtos concorrentes da marca A e marca B. Testar a hipótese de que a diferença das médias no índice de vendas entre as marcas é zero. Sendo $\alpha = 5\%$.

Supermercado	*Marca A*	*Marca B*
1	14	4
2	20	16
3	2	28
4	11	9
5	5	31
6	12	10

19. Com a finalidade de se estudar a influência da dieta no ganho de peso de prematuros, um grupo de 25 recém-nascidos (com pesos entre 1500 e 2000 g) foi dividido em 5 grupos de 5 crianças cada, e submetidas a diferentes dietas. Os dados abaixo representam o ganho médio diário em gramas:

	DIETA				
	A	*B*	*C*	*D*	*E*
	22	22	42	21	26
	31	26	30	21	19
	31	24	28	17	23
	26	21	26	19	25
	27	40	25	28	17
Σx_i	137	133	151	106	110
Σx_i^2	3811	3777	4749	2316	2480

a) Tome 2 dietas quaisquer e teste a igualdade de médias. $\alpha = 0{,}05$.

b) Repita o teste para outras duas.

20. Da população feminina extraiu-se uma amostra resultando:

Rendas (em $1000)	10 ⊢ 25	25 ⊢ 40	40 ⊢ 55	55 ⊢ 70	70 ⊢ 85
Nº de mulheres	7	12	10	6	4

Da população masculina retirou uma amostra resultando:

Rendas (em $1000)	15 ⊢ 30	30 ⊢ 45	45 ⊢ 60	60 ⊢ 75	75 ⊢ 90
Nº de Homens	8	15	12	7	3

Testar ao nível de 10% a hipótese de que a diferença entre a renda média dos homens e das mulheres valha $ 5000,00.

Testes para igualdade de duas proporções

21. Uma empresa de pesquisa de opinião pública seleciona, aleatoriamente, 300 eleitores de São Paulo e 400 do Rio de Janeiro, e pergunta a cada um se votará ou não no candidato *A* nas próximas eleições. 75 eleitores de SP e 120 do RJ respondem afirmativamente. Há diferença significativa entre as proporções de eleitores a favor de *A* naqueles dois Estados? Use $\alpha = 5\%$.

22. Estão em teste dois processos para fechar latas de comestíveis. Numa seqüência de 1.000 latas, o processo 1 gera 50 rejeições, enquanto o processo 2 acusa apenas 200 rejeições. Pode-se, no nível de 0,05, concluir que os dois processos sejam diferentes?

23. Numa pesquisa sobre possuidores de videocassete, encontram-se 120 das 200 casas pesquisadas do bairro *X* e 240 das 500 residências do bairro *Y*. Há diferença significativa entre a proporção de possuidores de vídeo nos dois bairros? Use $\alpha = 10\%$.

10

Estatística Não Paramétrica

10.1 INTRODUÇÃO

As técnicas da Estatística Não Paramétrica são, particularmente, adaptáveis aos dados das ciências do comportamento. A aplicação dessas técnicas não exige suposições quanto à distribuição da população da qual se tenha retirado amostras para análises. Podem ser aplicadas a dados que se disponham simplesmente em ordem, ou mesmo para estudo de variáveis nominais. Contrariamente ao que acontece na Estatística Paramétrica onde as variáveis são, na maioria das vezes, intervalares, como foi visto nos Capítulos 8 e 9. Os testes não-paramétricos são extremamente interessantes para análises de dados qualitativos.

Os testes da Estatística não paramétrica exigem poucos cálculos e são aplicáveis para análise de pequenas amostras ($n < 30$).

Como o próprio nome sugere, a estatística não-paramétrica independe dos parâmetros populacionais (μ; σ^2; σ; P ...) e de suas respectivas estimativas ($\bar{x}$, S^2, S, f...).

10.2 TESTE QUI-QUADRADO

O mais popular teste não-paramétrico é o teste qui-quadrado, ou teste de adequação do ajustamento.

Seja ε um experimento aleatório. Sejam E_1, E_2, ..., E_k, "K" eventos associados a ε. Admita que o experimento é realizado "n" vezes.

Sejam: Fo_1, Fo_2, ... Fo_k as freqüências observadas dos "K" eventos.

Sejam: Fe_1, Fe_2, ... Fe_k as freqüências esperadas, ou freqüências teóricas dos "K" eventos.

Deseja-se realizar um teste estatístico para verificar se há adequação de ajustamento entre as freqüências observadas e as freqüências esperadas. Isto é, se as discrepâncias $(Fo_i - Fe_i)$, $i = 1, 2, \ldots K$, são devidas ao acaso, ou se de fato existe diferença significativa entre as freqüências.

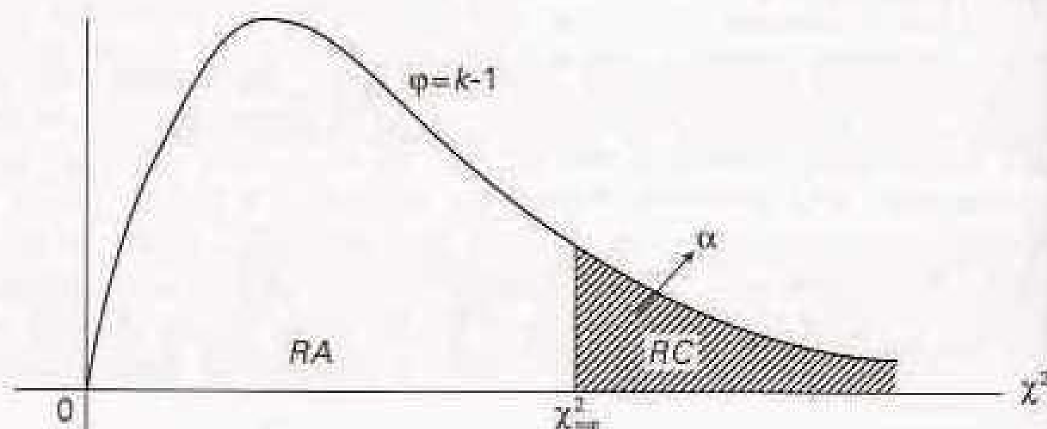
Eis o procedimento para se efetuar o teste.

1º) Enunciar as hipóteses H_0 e H_1.

H_0 afirmará não haver discrepância entre as freqüências observadas e esperadas, enquanto H_1 afirmará que as freqüências observadas e esperadas são discrepantes.

2º) Fixar α. Escolher a variável qui-quadrado com $\varphi = (K - 1)$. Lembre-se que K = número de eventos.

3º) Com auxílio da tabela χ^2, determinam-se RA e RC.



4º) Cálculo do valor da variável

$$\chi^2_{cal} = \sum_{i=1}^{K} \frac{(Fo_i - Fe_i)^2}{Fe_i} = \frac{(Fo_1 - Fe_1)^2}{Fe_1} + \ldots + \frac{(Fo_k - Fe_k)^2}{Fe_k}$$

5º) Conclusão

Se $\chi^2_{cal} < \chi^2_{sup}$, não se pode rejeitar H_0, ou seja, as freqüências observadas e esperadas não são discrepantes.

Se $\chi^2_{cal} > \chi_{sup}$, rejeita-se H_0, concluindo-se com risco α que há discrepância entre as freqüências observadas e esperadas. Ou seja, não há adequação do ajustamento.

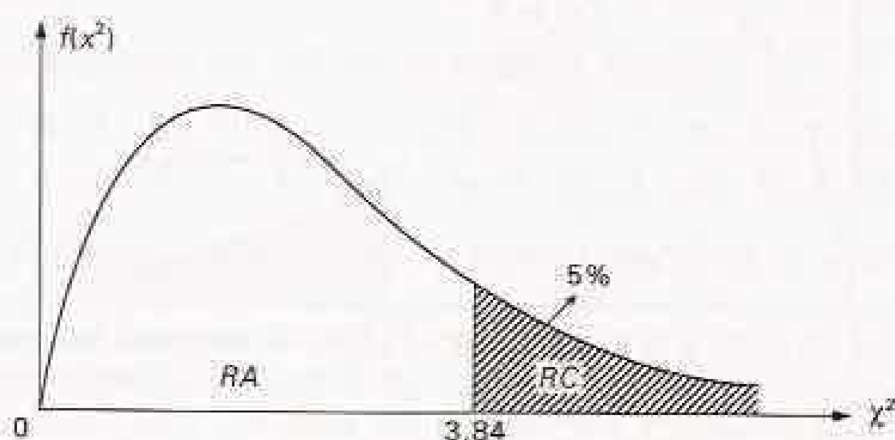
Exemplos:

1. Em 100 lances de uma moeda, observaram-se 65 coroas e 35 caras. Testar a hipótese de a moeda ser honesta, adotando-se $\alpha = 5\%$.

Solução: 1º) H_0: A moeda é honesta.
H_1: A moeda não é honesta.

2º) $\alpha = 5\%$. Escolhe-se uma χ_1^2, pois $K = 2$ e $\varphi = 2 - 1 = 1$.

3º) Determinação da RC e RA.



4º) Cálculo do valor da variável

Eventos	*Cara*	*Coroa*
Freqüências observadas	35	65
Freqüências esperadas	50	50

$$\chi_{cal}^2 = \sum_{i=1}^{2} \frac{(Fo_i - Fe_i)^2}{Fe_i} = \frac{(35-50)^2}{50} + \frac{(65-50)^2}{50} = 9$$

5º) Conclusão:
Como $\chi_{cal}^2 \geq 3{,}84$, rejeita-se H_0, concluindo-se, com risco de 5%, que a moeda não é honesta.

Eis outro exemplo:

Deseja-se testar se o número de acidentes numa rodovia se distribui igualmente pelos dias da semana. Para tanto foram levantados os seguintes dados:

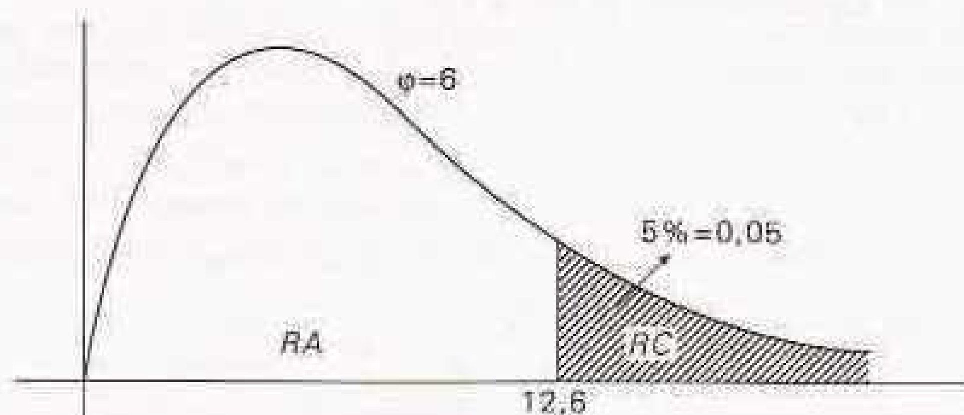
Dia da semana	Dom	Seg	Ter	Qua	Qui	Sex	Sáb
Número de acidentes	33	26	21	22	17	20	36

Adotar $\alpha = 5\%$.

Solução: 1. H_0 = As freqüências são iguais em todos os dias da semana.
H_1 = As freqüências são diferentes.

2. $\alpha = 5\%$. Escolhe-se uma variável qui-quadrado com $\varphi = K - 1 = 7 - 1 = 6$.

3. Determinação de RA e RC.



4. Cálculo do valor da variável.

Foes	33	26	21	22	17	20	36
Fesp	25	25	25	25	25	25	25

Observe: $\text{Fesp} = \frac{1}{7} \cdot 175 = 25$

$$\chi_{cal} = \frac{(33-25)^2}{25} + \frac{(26-25)^2}{25} + \frac{(21-25)^2}{25} + \frac{(22-25)^2}{25} +$$

$$+ \frac{(17-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(36-25)^2}{25} = 12{,}0$$

5. Conclusão

Como $\chi^2_{cal} < 12{,}6$, não se pode rejeitar H_0, logo as freqüências de acidentes não são diferentes para os dias da semana.

EXERCÍCIOS – SÉRIE 1 – CAPÍTULO 10

1. Uma moeda é lançada 200 vezes e verifica-se 110 caras e 90 coroas. Testar a honestidade da moeda, sendo $\alpha = 0{,}10$.

2. Um dado é lançado 180 vezes e as seguintes freqüências são observadas:

Eventos	Sair o 1	Sair o 2	Sair o 3	Sair o 4	Sair o 5	Sair o 6
Freqüências	31	28	35	26	29	31

Testar a hipótese do dado não ser viciado adotando $\alpha = 0{,}05$ e $\alpha = 0{,}025$.

3. Em 240 lances de um par de dados, observaram-se 17 somas "4" e 42 somas "7". Verificar se os dados são honestos adotando o nível de significância 0,01.

4. O último algarismo do CIC de 40 pessoas resultou:

2	8	0	4	5	7	3	7	7	4	3	2	1	0	9	6	5	9	1	9
8	0	3	3	2	1	9	8	7	6	6	0	1	2	4	9	3	7	6	4

Testar se é razoável supor esses números aleatórios, $\alpha = 5\%$.

5. O número de pessoas de certa raça que possua 4 tipos de sangue estaria nas proporções 0,18; 0,48; 0,20; 0,14. Dadas as freqüências observadas de 180; 360; 130 e 100 para uma outra raça, coloque à prova a hipótese da igualdade das distribuições. Adote $\alpha = 2{,}5\%$.

6. O número de livros emprestados por uma biblioteca, durante uma determinada semana, está indicado a seguir. Testar a hipótese de o número de livros emprestados não depender do dia da semana, sendo $\alpha = 0{,}01$.

Dias da semana	Seg	Ter	Qua	Qui	Sex
Número de livros emprestados	110	135	120	146	114

10.3 TESTE QUI-QUADRADO PARA INDEPENDÊNCIA OU ASSOCIAÇÃO

Uma importante aplicação do teste qui-quadrado ocorre quando se quer estudar a associação ou dependência, entre duas variáveis. A representação das freqüências observadas é dada por uma tabela de dupla entrada ou tabela de contingência.

O cálculo das freqüências esperadas fundamenta-se na definição de variáveis aleatórias independentes, conforme visto no Capítulo 2. Isto é: diz-se que X e Y são independentes se a distribuição conjunta de (X, Y) é igual ao produto das distribuições marginais de X e de Y. Isto é:

$P(x_i, y_j) = p(x_i) \cdot p(y_j)$ para todo i e j.

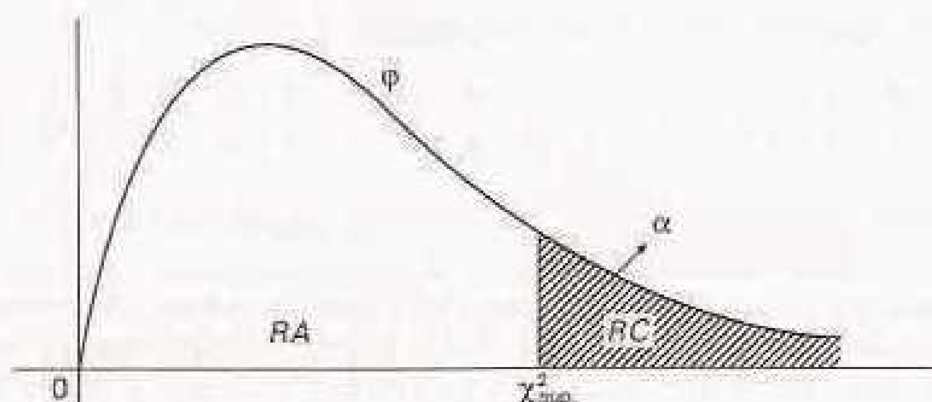
Procedimento para efetuar o teste:

1. H_0: as variáveis são independentes, ou as variáveis não estão associadas.

   H_1: as variáveis são dependentes, ou as variáveis estão associadas.

2. Fixar α. Escolher a variável qui-quadrado com $\varphi = (L-1)(C-1)$ onde L = número de linhas da tabela de contingência, e C = número de colunas.

3. Com auxílio da tabela χ^2, determinam-se RA e RC.



4. Cálculo do valor da variável.

$$\chi^2_{\text{cal}} = \sum_{i=1}^{L} \sum_{j=1}^{C} \frac{(Fo_{ij} - Fe_{ij})^2}{Fe_{ij}}$$

onde cada Fe_{ij} é determinado por

$$Fe_{ij} = \frac{(\text{soma da linha } i)(\text{soma da coluna } j)}{(\text{total de observações})}$$

5. Conclusão

   Se $\chi^2_{\text{cal}} < \chi^2_{\text{sup}}$, não se pode rejeitar H_0, isto é, não se pode dizer que as variáveis sejam dependentes.

   Se $\chi^2_{\text{cal}} > \chi^2_{\text{sup}}$, rejeita-se H_0, concluindo-se com risco α, que as variáveis são dependentes, ou estão associadas.

Exemplo: Testar ao nível de 5% se há dependência entre as preferências por sabor da pasta de dentes e o bairro.

Sabor da pasta	*Bairros*			Σ
	A	*B*	*C*	
limão	70	44	86	200
chocolate	50	30	45	125
hortelã	10	6	34	50
outros	20	20	85	125
Σ	150	100	250	500

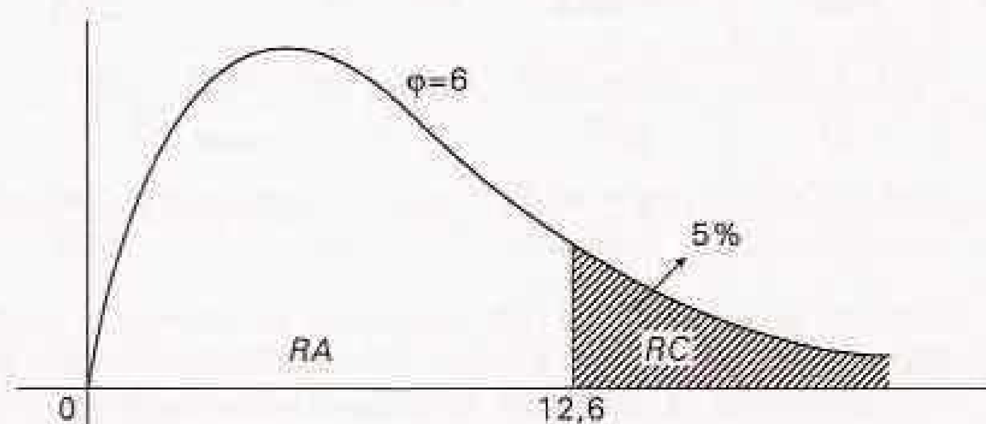
Solução: 1. H_0: A preferência pelo sabor independe do bairro.

H_1: A preferência pelo sabor depende do bairro.

2. $\alpha = 5\%$. Escolher um χ^2 com:

$$\varphi = (4-1)(3-1) = 6 \text{ gl}$$

3. RA e RC.



4. Cálculo do valor da variável.

A tabela das freqüências esperadas é dada por:

Sabor	*Bairros*		
	A (1)	*B (2)*	*C (3)*
(1) limão	60	40	100
(2) chocolate	37,5	25	62,5
(3) hortelã	15	10	25
(4) outros	37,5	25	62,5

onde, por exemplo,

$$Fe_{11} = \frac{(\text{soma da linha 1})\,(\text{soma da coluna 1})}{\text{total de observações}}$$

ou seja:

$$Fe_{11} = \frac{(150)\,(200)}{500} = 60$$

e

$$Fe_{43} = \frac{(\text{soma da linha 4})\,(\text{soma da coluna 3})}{\text{total de observações}}$$

$$Fe_{43} = \frac{(125)\,(250)}{500} = 62{,}5$$

Assim:

$$\chi^2_{cal} = \frac{(70-60)^2}{60} + \frac{(50-37{,}5)^2}{37{,}5} + \frac{(10-15)^2}{15} + \frac{(20-37{,}5)^6}{37{,}5} +$$

$$+ \frac{(44-40)^2}{40} + \frac{(30-25)^2}{25} + \frac{(6-10)^2}{10} + \frac{(20-25)^2}{25} +$$

$$+ \frac{(86-100)^2}{100} + \frac{(45-62{,}5)^2}{62{,}5} + \frac{(34-25)^2}{25} + \frac{(85-62{,}5)^2}{62{,}5} = 37{,}8\text{[illegible]}$$

5. Conclusão: Como $\chi^2_{cal} > 12{,}6$, rejeita-se H_0, concluindo-se, com risco de 5%, que há dependência entre sabor da pasta de dentes e o bairro.

Observações:

1. O teste qui-quadrado de associação é aconselhável quando o tamanho da amostra é razoavelmente grande e deve ser aplicado com maior cuidado se existem freqüências esperadas menores do que 5.

   Nestes casos, a solução é juntar classes adjacentes evitando-se que Fesp < 5.
2. Quando uma das variáveis possui níveis que contemplam todas as categorias da população (por exemplo: variável sexo: masculino e feminino, não sendo possível outra modalidade de sexo ...) diz-se que o teste é de homogeneidade.

3. Um indicador do grau de associação entre duas variáveis analisadas pelo teste qui-quadrado é o *coeficiente de contingência*, dado por:

$$C = \sqrt{\frac{\chi^2_{cal}}{\chi^2_{cal} + n}}$$

O coeficiente pode variar entre [0,1], estando mais associadas as variáveis quanto maior é o valor de C. O limite superior do coeficiente depende do tamanho da tabela de contingência. Assim, para tabelas (2 × 2), o máximo de C é dado por 0,7071. Sendo maior para tabelas maiores.

EXERCÍCIOS – SÉRIE II – CAPÍTULO 10

1. Testar (α = 5%) se há alguma relação entre as notas escolares e o salário.

Notas Escolares.

SALÁRIO		***Alta***	***Média***	***Baixa***
	Alto	18	17	5
	Médio	26	38	16
	Baixo	6	15	9

2. A tabela apresenta os resultados de um experimento destinado a investigar o efeito da vacinação de animais contra determinada doença. Testar a homogeneidade dos resultados utilizando: a) α = 5%; b) α = 1%.

	Contraíram a doença	***Não contraíram a doença***
Vacinados	14	42
Não vacinados	16	28

3. Determine o valor do coeficiente de contingência considerando os dados:

	Arena	***PMDB***
Homens	50	72
Mulheres	29	35

4. Verificar se há associação entre os níveis de renda e os municípios onde foram pesquisados 400 moradores. Adote α = 1%.

Municípios	Níveis de renda			
	A	B	C	D
1	28	42	30	24
2	44	78	78	76

5. Os dados abaixo são o resultado de um questionário aplicado a 500 eleitores

Opinião a respeito da pena de morte	Partidos		
	Esquerda	Centro	Direita
Aprovam	35	50	80
Não aprovam	45	80	60
Sem opinião	20	70	60

Os dados sugerem que a opinião em relação à pena de morte seja independente do partido? Use $\alpha = 5\%$.

10.4 TESTE DOS SINAIS

É utilizado para análise de dados emparelhados (o mesmo indivíduo é submetido a duas medidas). É aplicado em situações em que o pesquisador deseja determinar se duas condições são diferentes.

A variável de estudo poderá ser intervalar ou ordinal. O nome "teste dos sinais" se deve ao fato de se utilizar em sinais "mais" e "menos" em lugar dos dados numéricos. Assim, se houve alteração para maior, usa-se (+), se para menor, (–). Não havendo alteração, atribui-se (0). Para o teste desconsideram-se os casos de empate, ou seja, os pares em que foram atribuídos zeros.

A "lógica" do teste é que as condições podem ser consideradas iguais quando as quantidades de "+" e "–" forem aproximadamente iguais. Isto é, a proporção de sinais "+" equivale a 50%, ou seja: $p = 0,5$.

Procedimento:

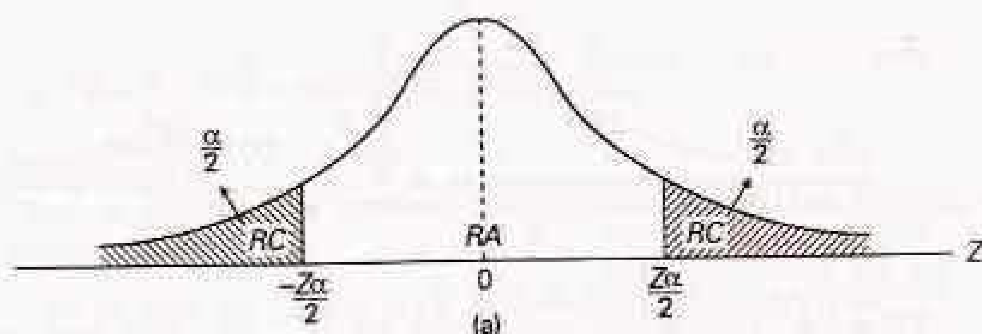
1. H_0: não há diferença entre os grupos, ou seja: $p = 0,5$.

   H_1: há diferença, ou seja: uma das alternativas

   $p \neq 0,5$ (*a*)
   $p < 0,5$ (*b*)
   $p > 0,5$ (*c*)

2. Fixar α. Escolher a distribuição $N(0,1)$ se $n > 25$ ou Binomial se $n \leq 25$.

3. Com auxílio da tabela, determinam-se RA e RC (para $n > 25$), caso $n < 25$ utiliza-se a distribuição Binomial.



(a)

4. Cálculo do valor da variável ($n > 25$)

$$Z_{cal} = \frac{y - n.p.}{\sqrt{n.p.q.}}$$

onde:

y = número de sinais "+"
n = tamanho da amostra descontados os empates
p = 0,5
q = $1 - p = 0,5$

5. Conclusões:

Se $-Z\frac{\alpha}{2}\ Z_{cal} \leq Z\frac{\alpha}{2}$, não se pode rejeitar H_0.

Se $Z_{cal} > Z\frac{\alpha}{2}$ ou $Z_{cal} < -Z\frac{\alpha}{2}$, rejeita-se H_0, concluindo-se, com risco α, que há diferença entre os dois grupos, ou duas condições.

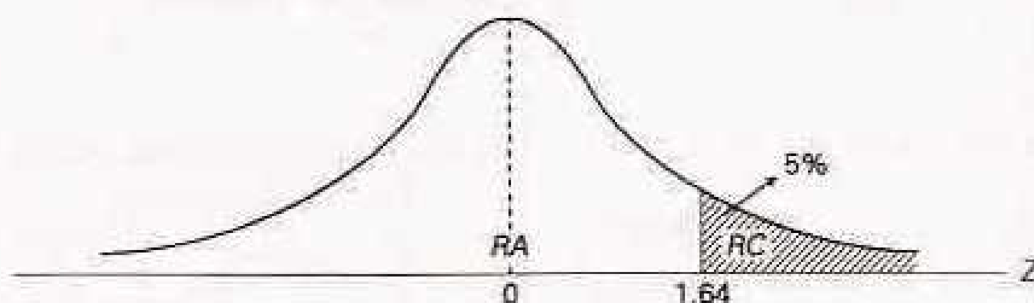
Exemplo: Sessenta alunos matriculam-se num curso de inglês. Na primeira aula aplica-se um teste que mede o conhecimento da língua. Após seis meses, aplica-se um segundo teste. Os resultados mostram que 35 alunos apresentaram melhora (35 "+"), 20 se conduziram melhor no primeiro teste (20 "–") e 5 não apresentaram modificação (5 "0"). Testar, no nível de 5%, se o curso alterou o conhecimento de inglês do grupo de 60 alunos.

Solução: 1. H_0: O curso não alterou ($p = 0,5$).

H_1: O curso melhorou o conhecimento de inglês.

2. α = 5%. Variável $N(0,1)$

3. RA e RC



Observe, devido ao enunciado de H_1, optou-se pelo teste unicaudal à direita. Caso H_1 fosse "piorou", o teste seria unicaudal à esquerda.

4. Cálculo do valor da variável.

$$Z_{cal} = \frac{35 - 55\,(0,5)}{\sqrt{55\,(0,5)\,(0,5)}} = 2,02$$

onde: $y = 35$

$n = 60 - 5 = 55$

$p = q = 0,5$

5. Conclusão

Como $Z_{cal} > 1,64$, rejeita-se H_0, concluindo-se, com risco de 5%, que o curso melhorou o conhecimento de inglês.

10.5 TESTE DE WILCOXON

Trata-se de uma extensão do teste dos sinais. É mais interessante do que aquele, pois leva em consideração a magnitude da diferença para cada par.

Procedimento:

Antes de enunciar as hipóteses é preciso:

a) Determinar para cada par a diferença (*di*) entre os dois escores.
b) Atribuir postos (colocar em ordem crescente) a todos os "di"s, desconsiderando-se os sinais. No caso de empate, atribuir a média dos postos empatados.

c) Identificar cada posto pelo sinal "+" ou "–" do "di" que ele representa.

d) Determinar T = a menor das somas de postos de mesmo sinal.

e) Abater do "n" o número de zeros, isto é, $di = 0$.

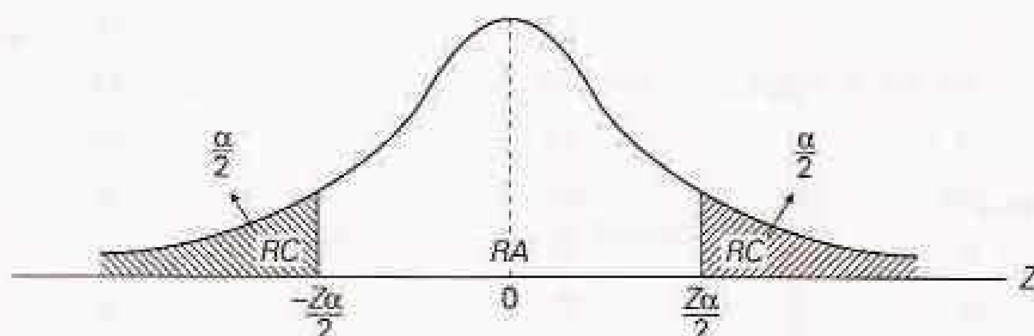
Eis o teste:

1. H_0: não há diferença entre os grupos.

   H_1: há diferença.

2. Fixar α. Escolher a variável $N(0,1)$ se $n > 25$. Se $n \leq 25$ há tabela própria.

3. Com auxílio da tabela $N(0,1)$ determinam-se RA e RC.



4. Cálculo do valor da variável

$$Z_{cal} = \frac{T - \mu_T}{\sigma_T}$$

onde: T = menor das somas de postos de mesmo sinal

$$\mu_{(T)} = \frac{n(n+1)}{4}$$

$$\sigma_{(T)} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

5. Conclusão:

   Se $-Z\frac{\alpha}{2} \leq Z_{cal} \leq Z\frac{\alpha}{2}$, não se pode rejeitar H_0.

   Se $Z_{cal} > Z\frac{\alpha}{2}$ ou $Z_{cal} < -Z\frac{\alpha}{2}$, rejeita-se H_0, concluindo-se, com risco α, que há diferença entre os dois grupos, ou duas condições.

Exemplo: Será apresentado um exemplo com $n < 25$, todavia, não se utilizará a tabela apropriada. O propósito é ilustrar o procedimento do teste.

Um grupo de 15 pessoas submeteu-se a um novo processo de emagrecimento. Testar no nível de 5% a hipótese de que não houve diminuição significativa do peso.

Indivíduos	*Peso (kg) antes da dieta*	*Peso (kg) depois da dieta*
1	85	83
2	78	70
3	65	67
4	80	80
5	58	60
6	58	55
7	93	80
8	80	79
9	80	84
10	65	60
11	62	60
12	87	87
13	95	90
14	58	57
15	90	94

Solução: Convém apresentar os dados numa tabela do tipo:

Peso-antes	*Peso-depois*	*di*	*Postos (+)*	*Postos (–)*
85	83	(85 – 83) = 2	4,5º	
78	70	8	12º	
65	67	–2		4,5º
60	80	0		
58	60	–2		4,5º
58	55	3	7º	
93	80	13	13º	

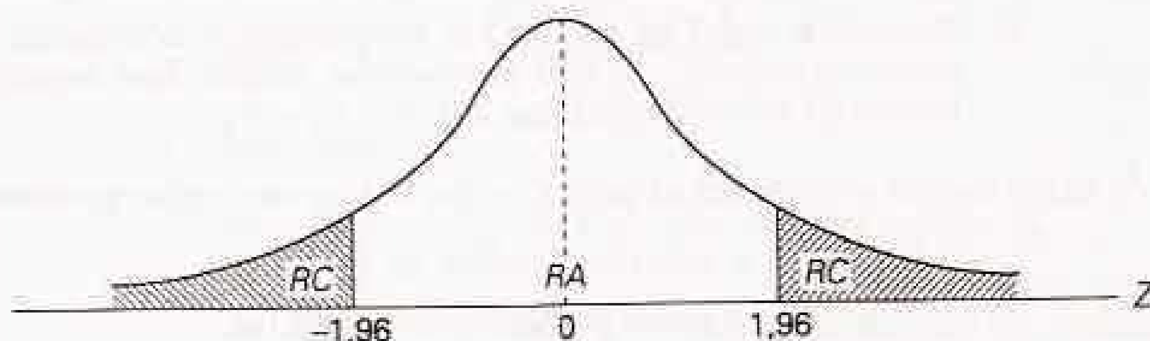
continua

Peso-antes	Peso-depois	di	Postos (+)	Postos (–)
80	79	1	1,5º	
80	84	–4		8,5º
65	60	5	10,5º	
62	60	2	4,5º	
87	87	0		
95	90	5	10,5º	
58	57	1	1,5º	
90	94	–4		8,5º
		Σ	65	26

Veja:

- houve empate para *di* = 1, então o posto atribuído foi:
  $\frac{1^{\circ} + 2^{\circ}}{2} = 1{,}5^{\circ}$
- para *di* = 2, também houve empate, assim, $\frac{3^{\circ} + 4^{\circ} + 5^{\circ} + 6^{\circ}}{4} = 4{,}5^{\circ}$
- será o posto atribuído ao "2"
  e dessa maneira para os demais casos.

Teste: 1. H_0: não houve diferença dos pesos.
H_1: houve diferença.
2. α = 5%. Escolher a distribuição N (0,1).
3. Determinação de RA e RC.



4. Cálculo do valor da variável.

$T = 10$ (menor das somas)

Lembre-se: $n = 15$, mas houve dois empates, logo

$$n^* = 13 \qquad \mu_T = \frac{13(13+1)}{4} = 45{,}5$$

$$\sigma_{(T)} = \sqrt{\frac{13(13+1)(2(13)+1)}{24}} = 14{,}31$$

$$Z_{cal} = \frac{26 - 45{,}5}{14{,}31} = -1{,}36$$

5. Como $-1{,}96 \le Z_{cal} \le 1{,}96$, não se pode rejeitar H_0 ao nível de 5%.

10.6 TESTE DE MANN-WHITNEY

É usado para testar se duas amostras independentes foram retiradas de populações com médias iguais. Trata-se de uma interessante alternativa ao teste paramétrico para igualdade de médias, pois o teste Mann-Whitney não exige nenhuma consideração sobre as distribuições populacionais e suas variâncias. Como foi visto, o teste paramétrico para igualdade de médias exige populações com distribuições normais de mesma variância.

Este teste poderá ser aplicado para variáveis intervalares ou ordinais.

Procedimento:

a) Considerar n_1 = número de casos do grupo com menor quantidade de observações e n_2 = número de casos do maior grupo.

b) Considere todos os dados dos dois grupos e coloque-os em ordem crescente. Atribua primeiro ao escore que algebricamente for menor e prossiga até $N = n_1 + n_2$.

Às observações empatadas atribuir a média dos postos correspondentes:

c) Calcular R_1 = soma dos postos do grupo n_1.

R_2 = soma dos postos do grupo n_2.

d) Escolher a menor soma entre R_1 e R_2.

e) Calcular a estatística:

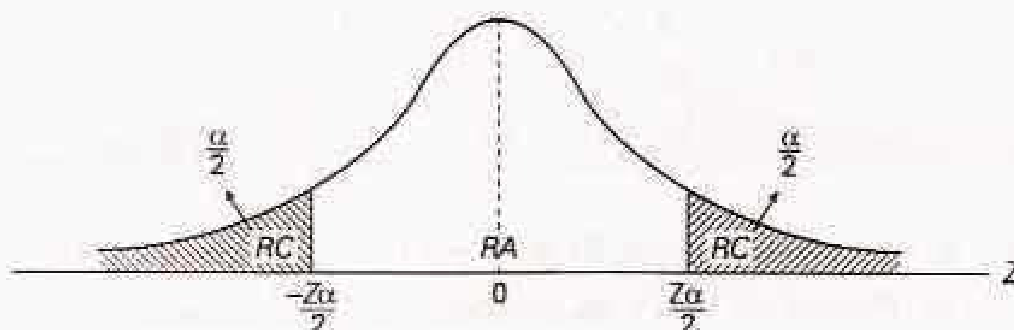
$$\mu_1 = n_1 \cdot n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

ou

$$\mu_2 = n_1 \cdot n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

Teste:

1. H_0: não há diferença entre os grupos.
   H_1: há diferença.
   Para n_1 , $n_2 < 10$ há tabela própria.
2. Fixar α. Escolher a variável $N(0,1)$.
3. Com auxílio da tabela $N(0,1)$ determinam-se RA e RC.



4. Cálculo do valor da variável.

$$Z_{cal} = \frac{\mu - \mu(u)}{\sigma(u)}$$

onde:

$$\mu(u) = \frac{n_1 \cdot n_2}{2}$$

$$\sigma(u) = \sqrt{\frac{n_1 \cdot n_2(n_1 + n_2 + 1)}{12}}$$

5. Conclusão:

   Se $-Z\frac{\alpha}{2} \leq Z_{cal} \leq Z\frac{\alpha}{2}$, não se pode rejeitar H_0.

   Se $Z_{cal} > Z\frac{\alpha}{2}$ ou $Z_{cal} < -Z\frac{\alpha}{2}$, rejeita-se H_0, concluindo, com risco α, que há diferença entre os grupos.

Exemplo: Determine, no nível de 10%, se as vendas médias dos dois "shop-pings" são diferentes.

Shopping A	*Shopping B*
(em 10^6 $)	
10	22
18	17
9	15
8	10
2	7
11	7
4	8
3	14
9	15
12	
10	

Solução: a) $n_1 = 9$ (shopping B) e $n_2 = 11$.

b) Postos de todas as vendas.

	A	*B*
	11º	20º
	19º	18º
	8,5º	16,5º
	6,5º	11º
	1º	4,5º
	13º	4,5º
	3º	6,5º
	2º	15º
	8,5º	16,5º
	14º	
	11º	
Soma	97,5	112,5

$R_2 = 97,5$ $R_1 = 112,5$

d) Escolher $R_2 = 97,5$

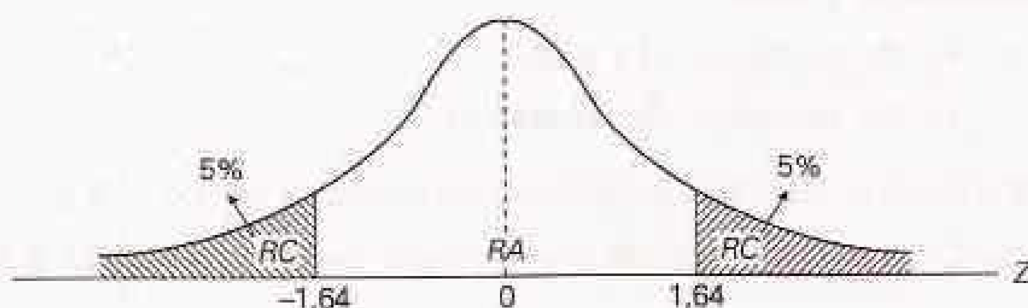
e) $u_2 = 9(11) + \dfrac{11(11+1)}{2} - 97,5 = 67,5$

Teste:

1. H_0: as vendas são iguais.
   H_1: as vendas são diferentes.

2. $\alpha = 10\%$. Escolher $N(0,1)$

3. Com auxílio da tabela:



4. Cálculo do valor da variável

$$\mu(u) = \frac{(9)(11)}{2} = 49,5$$

$$\sigma(u) = \sqrt{\frac{9(11)(9 + 11 + 1)}{12}} = 13,16$$

$$Z_{cal} = \frac{67,5 - 49,5}{13,16} = 1,37$$

5. Conclusão:
   Como $-1,64 \leq Z_{cal} \leq 1,64$, não se pode rejeitar a hipótese de que as vendas são iguais.

10.7 TESTE DA MEDIANA

Trata-se de uma alternativa ao teste de Mann-Whitney. Testa-se a hipótese de dois grupos independentes tenham medianas iguais. É aplicado para variáveis ordinais ou intervalares.

Procedimento:

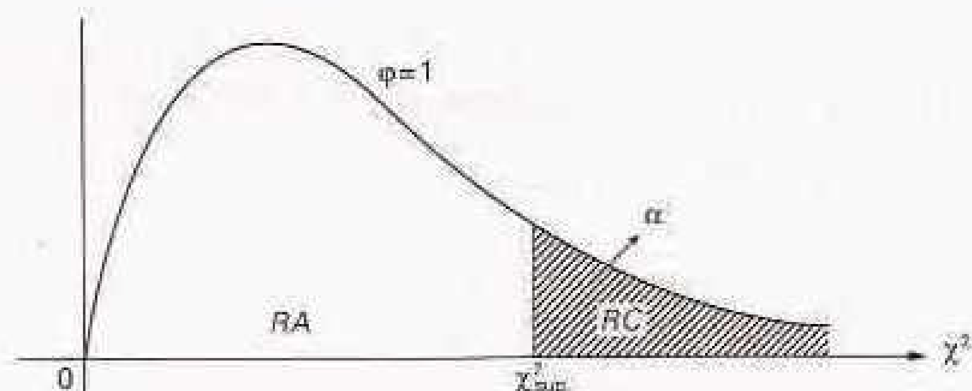
a) Determina-se a mediana do grupo combinado (duas amostras juntas

b) Monta-se a tabela:

	Grupo I		Grupo II	
Freqüências observadas acima da mediana	Fobs	Fesp	Fobs	Fesp
Freqüências observadas abaixo ou igual à mediana	Fobs	Fesp	Fobs	Fesp

c) Realiza-se o teste:

1. H_0: As medianas são iguais.
   H_1: As medianas são diferentes.
2. Fixar α. Escolher a variável qui-quadrado com $\varphi = 1$ gl.
3. Com auxílio da tabela qui-quadrado determinam-se RA e RC.



4. Cálculo do valor da variável.

$$\chi^2_{cal} = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(Fo_{ij} - Fe_{ij})^2}{Fe_{ij}}$$

As freqüências observadas são obtidas pela contagem dos elementos de cada grupo que se encontram acima da mediana, e aqueles que estão abaixo ou são iguais à mediana.

As freqüências esperadas são obtidas da mesma maneira que aquela utilizada para o teste χ^2 de independência:

$$Fe_{ij} = \frac{(\text{soma da linha } i)\ (\text{soma da coluna } j)}{(\text{total de observações})}$$

5. Conclusão:

Se $\chi^2_{cal} \leq \chi^2_{sup}$, não se pode rejeitar H_O.

Se $\chi^2_{cal} > \chi^2_{sup}$, rejeita-se H_0, concluindo-se com risco α que as medianas são diferentes.

Exemplo: Duas turmas de Estatística de duas faculdades distintas obtiveram as seguintes notas:

Fac. *X*: 8 – 7 – 5 – 5 – 10 – 4 – 6 – 9 – 3

Fac. *Y*: 1– 2 – 4 – 6 – 8 – 10 – 7

Testar no nível de 5% a igualdade das medianas.

Solução: Calcula-se a mediana do grupo combinado.

X_i	F_i	F_{ac}	
1	1	1	
2	1	2	
3	1	3	
4	2	5	
5	2	7	
6	2	9	→ 8º e 9º
7	2	11	
8	2	13	
9	1	14	
10	2	16	
Σ	16		

$$\frac{n}{2} = \frac{16}{2} = 8^{\underline{o}}$$

$$\frac{n}{2} + 1 = \frac{16}{2} + 1 = 9^{\underline{o}}$$

$$\text{logo } Md = \frac{6 + 6}{2} = 6$$

Monta-se a tabela:

	Fac. X	Fac. Y	Σ
Freqüências acima da mediana	4* / 3,94	3 / 3,06	7
Freqüências abaixo ou igual à mediana	5 / 5,06	4 / 3,94	9
	9	7	16

Observe que $Fe_{11} = \frac{9 \times 7}{16} = 3,94$

* Trata-se da freqüência observada de notas acima de 6 na Faculdade X. Confira.

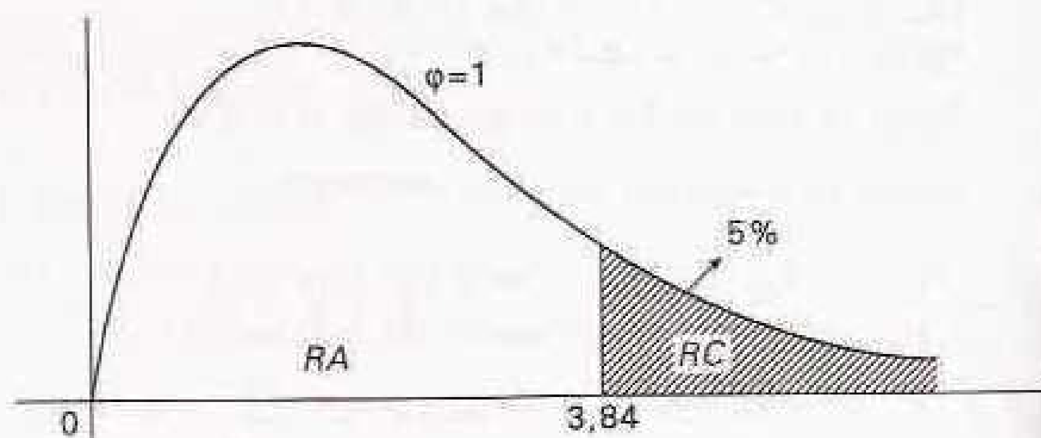
O teste:

1. H_0: As medianas são iguais:

   H_1: As medianas são diferentes.

2. $\alpha = 5\%$. Escolhe-se a variável qui-quadrado com 1 gl.

3. RA e RC.



4. Cálculo do valor da variável (*)

$$\chi^2_{cal} = \frac{(4-4)^2}{4} + \frac{(5-5)^2}{5} + \frac{(3-3)^2}{3} + \frac{(4-4)^2}{4} = 0$$

5. Conclusão: Como $\chi^2_{cal} < 3{,}84$, não se pode rejeitar a hipóte[se] de que as medianas são iguais.

10.8 TESTE KRUSKAL-WALLIS

Trata-se de teste extremamente útil para decidir se K amostras ($K >$ 2) independentes provêm de populações com médias iguais. Poderá ser aplica[do] para variáveis intervalares ou ordinais.

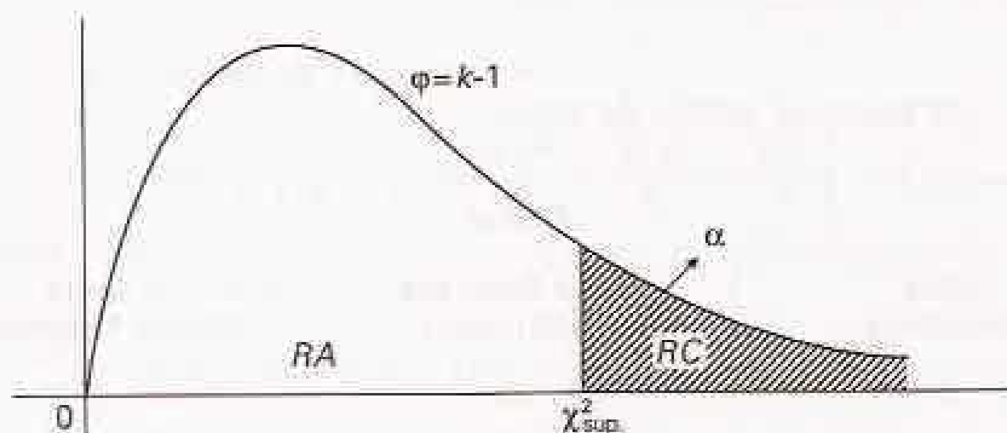
Procedimento:

a) Dispor, em ordem crescente, as observações de todos os K grupos, atribuindo-lhes postos de 1 a n. Caso haja empates, atribuir o posto médio.

b) Determinar o valor da soma dos postos para cada um dos K grupos: R_i $i = 1, 2, ..., K$.

c) Realizar o teste:

1. H_0: As médias são iguais.
   H_1: Há pelo menos um par diferente.
2. Fixar α. Escolher uma variável qui-quadrado com $\varphi = k - 1$.
3. Com auxílio da tabela qui-quadrado determinam-se RA e RC.



4. Calcula-se a estatística:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{K} \frac{(R_i)^2}{n_i} - 3(n+1)$$

5. Conclusão:
   Se $H < \chi^2_{sup}$, não se pode rejeitar H_0.
   Se $H > \chi^2_{sup}$, rejeita-se H_0 concluindo-se com risco α que há diferença entre as médias dos K grupos.

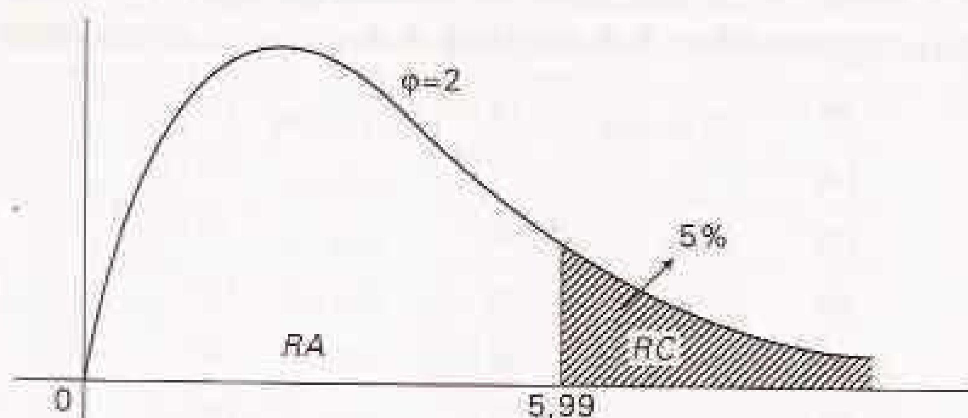
Exemplo: Testar, no nível de 5%, a hipótese da igualdade das médias para os três grupos de alunos que foram, submetidos a esquemas diferenciados de aulas. Foram registradas as notas obtidas para uma mesma prova.

Aulas Expositivas	Aulas com recursos audiovisuais	Aulas através de ensino programado
65	60	61
62	71	69
68	66	67
70	63	72
60	64	74
	59	

Solução: Atribuem-se postos às notas:

Postos		
Aulas Expositivas	**Aulas Recursos Audiovisuais**	**Aulas Ensino Programado**
8º	2,5º	4º
5º	14º	12º
11º	9º	10º
13º	6º	15º
2,5º	7º	16º
	1º	
Σ 39,5º	39,5º	57º

1. H_0: as notas médias são iguais.
   H_1: as notas médias são diferentes.
2. $\alpha = 5\%$. Escolhe-se uma distribuição qui-quadrado com 2 g.l. pois $\varphi = K - 1 = 3 - 1 = 2$.
3. Com auxílio da tabela da distribuição determinam-se RA e RC.



4. Cálculo da estatística *H*.

$$H = \frac{12}{16\,(16+1)}\left[\frac{39{,}5^2}{5} + \frac{39{,}5^2}{6} + \frac{57^2}{5}\right] - 3\,(16+1)$$

$H = 2{,}90$.

5. Conclusão: Como $H < 5{,}99$, não se pode rejeitar H_0.

   Assim, as notas médias podem ser consideradas iguais, ao nível de 5%.

EXERCÍCIOS – SÉRIE III – CAPÍTULO 10

1. Para cada uma das situações aplique o teste dos sinais e o teste de Wilcoxon. Adote $\alpha = 2{,}5\%$.

   a) Indivíduos submetidos a um programa de dieta.

Peso (kg) pré-dieta	Peso (kg) pós-dieta	Continuação	
55	50	48	50
63	65	49	51
78	78	90	81
81	79	93	85
68	70	90	90
58	57	56	58
60	58	66	64
60	62	67	68

Peso (kg) pré-dieta	*Peso (kg) pós-dieta*	*Continuação*	
75	70	73	70
85	81	74	70
90	80	48	53
50	60	68	65
58	55	72	70
83	75	86	83
47	52	80	81

b) Veículos com um novo aditivo.

Km/l			
Antes	*Depois*	*Antes*	*Depois*
8,7	9,3	4,8	5,5
9,8	9,2	6,7	6,8
10,0	9,5	8,3	8,5
9,6	9,6	9,5	9,0
8,5	8,8	10,5	10,0
5,8	6,5	12,5	13,0
6,3	7,0	12,5	12,0
12,5	11,5	9,0	10,0
8,8	8,9	14,0	12,0
7,3	8,0	13,0	11,0
12,5	11,0	9,5	10,5
13,8	14,0	8,0	9,3

2. Use o teste de Mann-Whitney para determinar se a média do grupo X é maior do que a média do grupo Y. Adote $\alpha = 1\%$.

X: 63 65 70 48 50 81 88 99 35 47 75 85 61

Y: 90 50 60 70 40 38 89 47 51 65 87

3. Aplique o teste da mediana nos dados do Exercício 2. Adote $\alpha = 1\%$. Compare as conclusões.

4. São dadas as durações (em km) de duas marcas de amortecedores. Aplique um teste para verificar a igualdade das quilometragens. Use α = 5%.

Marca A	*Marca B*
26.560	33.400
21.900	29.600
28.800	25.500
27.700	27.900
31.800	24.500
24.500	23.800
27.800	27.800
30.600	30.100
25.600	28.860
24.600	27.700
25.400	24.450
35.500	32.300
26.300	34.300
27.900	
28.400	

5. Testaram-se quatro tipos de lâmpadas para determinar se havia diferença entre suas vidas médias. Adote α = 5% para realizar o teste estatístico que confirma, ou não, a igualdade da duração.

MARCA A	704	604	1.038	881	924	672	723	591
MARCA B	752	709	717	921	761	991	805	981
MARCA C	873	666	1.021	992	816	918	978	1.203
MARCA D	690	850	824	856	915	734	799	700

6. Verificar se há diferença entre as vendas médias dos shoppings. Use α = 2,5%.

(Em $ 1.000.000)								
A	3,2	4,8	5,0	2,7	1,8	6,0	7,0	5,5
B	6,2	1,3	1,7	2,0	5,0	2,3		
C	5,0	4,0	3,0	2,0	1,.7	1,0	4,5	

11

Comparação de Várias Médias – Análise da Variância

11.1 INTRODUÇÃO

Nos capítulos anteriores foram apresentados testes, paramétrico e não-paramétrico, para verificar a igualdade entre duas médias: teste "t" e de Mann-Whitney. Ainda no Capítulo 10 o teste de Kruskal-Wallis foi aplicado para testar a igualdade de K médias, $K > 2$.

Uma alternativa ao teste de Kurskal-Wallis é a análise da variância.

Trata-se de um método estatístico, desenvolvido por Fisher, que através de testes de igualdades de médias, verifica se fatores produzem mudanças sistemáticas em alguma variável de interesse. Os fatores propostos podem ser variáveis quantitativas ou qualitativas, enquanto a variável dependente deve ser quantitativa (intervalar) e é observada dentro das classes dos fatores – os tratamentos.

Por exemplo, pode-se estar interessado no consumo de combustível dos automóveis. Poder-se-ia admitir que a marca do veículo, idade, potência etc. como fatores. Por meio da análise da variância é possível verificar se as marcas, idade, potência ...., ou uma combinação desses fatores produzem efeitos apreciáveis sobre o consumo, ou se concluir que tais fatores não têm influência sobre o consumo.

A finalidade é apresentar os fundamentos esse método. Para estudos mais aprofundados recomenda-se consultas a livros que tratam, com exclusividade, desse assunto.

11.2 HIPÓTESES DO MODELO

Há três suposições básicas que devem ser satisfeitas para que se possa aplicar a análise da variância.

1. As amostras devem ser aleatórias e independentes.
2. As amostras devem ser extraídas de populações normais.
3. As populações devem ter variâncias iguais.

11.3 CLASSIFICAÇÃO ÚNICA OU EXPERIMENTO DE UM FATOR

Admite-se um único fator (variável independente) que é subdividido em tratamentos (níveis do fator). A variável de estudo (variável dependente) é medida através de amostras de cada tratamento. Eis a configuração desse tipo de experimento:

Tratamentos / *Elemento da amostra*	1	2	3	....	K	
1	x_{11}	x_{21}	x_{31}	....	x_{K1}	
2	x_{12}	x_{22}	x_{32}	....	x_{K2}	
3	x_{13}	x_{23}	x_{33}	....	x_{K3}	
.	.	.	.		.	
.	.	.	.		.	
.	.	.	.		.	
n_i	x_{1n1}	x_{2n2}	x_{3n3}	....	x_{KnK}	
Somas						Total
Médias	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	....	$\bar{x}_K$	$\bar{X}$

$i = 1, 2, 3, ..., K$
$j = 1, 2, 3, ..., n_i$

Assim x_{ij} denota o valor da j-ésima observação sujeito ao i-ésimo tratamento.

A média dos valores observados no *i*-ésimo grupo será

$$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} \qquad i = 1, 2, \ldots, K$$

A média geral é dada por:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij} \quad \text{ou} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{K} n_i \bar{x}_i$$

em que $n = \sum_{i=1}^{K} n_i$ é o número total de observações.

A hipótese nula é de que todos os tratamentos tenham médias iguais, isto é:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_K$$

E que todas as "K" populações dos tratamentos tenham a mesma variância: σ^2.

A hipótese alternativa é de que pelo menos um par de médias seja diferente:

$$H_1: \mu_p \neq \mu_q \quad \text{para} \quad p \neq q$$

A aceitação de H_0 revelará que o fator considerado não acarreta mudanças significativas na variável de estudo. Por outro lado, a rejeição de H_0 indicará , com risco α, que o fator considerado exerce influência sobre a variável de estudo.

A base da análise da variância está nas comparações que podem ser feitas com os estimadores da variância comum de todos os tratamento (σ^2).

11.3.1 Estimadores da Variância Comum σ^2

Admitindo-se H_0 como verdadeira, pode-se estimar a variância comum de três maneiras diferentes:

1ª) No primeiro caso consideram-se os K tratamentos como uma única amostra de tamanho n e a média geral $\bar{X}$. Se H_0: $\mu_1 = \mu_2 = \ldots = \mu_K = \mu$ é verdadeira, tem-se que:

$$S_t^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2}{n - 1}$$

será um estimador justo de σ^2, isto é, $E[S_t^2] = \sigma^2$. Por outro lado, se H_0 não for verdadeira, S_t^2 irá superestimar σ^2 (veja configuração à frente).

Ao numerador $\sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2$ denomina-se variação total (Qt).

Pelo Teorema de Fisher: $\frac{\Sigma\Sigma (x_{ij} - \bar{x})^2}{\sigma^2}$, tem distribuição qui-quadrado com $(n - 1)$ graus de liberdade.

Desenvolvendo-se o quadrado obtém-se uma fórmula prática para o cálculo da variação total. Assim:

$$\text{Variação total} = Qt = \sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij}^2 - C$$

em que

$$C = \frac{\left(\sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij} \right)^2}{n}$$

2ª) A segunda forma de se estimar a variância comum σ^2 é pela consideração das médias dos K grupos e a média geral $\bar{X}$. Se H_0 for verdadeira, teremos para cada amostra:

$$E[\bar{x}_i] = \mu \quad \text{e} \quad \sigma^2(\bar{x}_i) = \frac{\sigma^2}{n_i}, \quad \text{ou} \quad \bar{x}_i \overset{d}{=} N\left(\mu; \frac{\sigma^2}{n_i}\right). \text{ Então}$$

$$S^2 = \frac{\sum_{i=1}^{K} (\bar{x}_i - \bar{x})^2}{K-1} \text{ será um estimador justo de } \frac{\sigma^2}{n_i} \text{ e}$$

$$\boxed{S_e^2 = n_i S^2 = \frac{\sum_{i=1}^{K} n_i (\bar{x}_i - \bar{x})^2}{K-1}}$$

será um estimador justo de σ^2, isto é: $E[S_e^2] = \sigma^2$.

Porém, se H_0 não for verdadeira, S_e^2 irá superestimar σ^2. (Veja a configuração à frente).

Ao numerador $\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x})^2$ denomina-se variação entre tratamentos.

Pelo teorema de Fisher:

$$\frac{\sum_{i=1}^{K} n_i (\bar{x}_i - \bar{x})^2}{\sigma^2}$$, tem distribuição qui-quadrado com $(k - 1)$ graus de liberdade.

Eis a fórmula prática para o cálculo de Q_e.

$$\boxed{Q_e = \sum_{i=1}^{K} \left[\frac{\left(\sum_{j} x_{ij} \right)^2}{n_i} \right] - C}$$

3ª) A terceira maneira de se estimar a variância σ^2 comum será por meio de cada uma das K amostras. Assim, para o i-ésimo tratamento, tem-se:

$$S_i^2 = \frac{\sum_{i=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n_i - 1}$$

Como $i = 1, 2, \ldots, K$ têm-se "K" estimadores do tipo S_i^2. Então o estimador da variância comum será dado pela média aritmética dos S_i^2 ponderadas pelos respectivos graus de liberdade ($\phi_i = n_i - 1$), assim:

$$S_r^2 = \frac{\phi_1 S_1^2 + \phi_2 S_2^2 + \ldots + \phi_K S_K^2}{\phi_1 + \phi_2 + \ldots + \phi_K}$$

ou seja

$$S_r^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n - K}$$

Sob a condição de H_0 ser verdadeira ou não, tem-se $E[S_r^2] = \sigma^2$, isto é, S_n^2 é justo. (Veja configuração à frente).

Ao numerador $\sum_{i=1}^{K} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ denomina-se variação dentro dos tratamentos ou variação residual (Q_r).

Pelo teorema de Fisher:

$\dfrac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{\sigma^2}$, tem distribuição qui-quadrado com $(n - k)$ graus de liberdade.

Eis a fórmula prática para o cálculo de Q_r.

$$Q_r = \sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i} \left[\frac{\left(\sum_{j} x_{ij} \right)^2}{n_i} \right]$$

Pode-se demonstrar que:

$$Q_t = Q_e + Q_r$$

Isto é, a variação total é igual à soma da variação entre tratamentos e a variação residual. Ou seja, $Q_t - Q_e - Q_r = 0$.

Resumindo, as variações Q_t, Q_e e Q_r têm distribuição χ^2 respectivamente com $(n - 1)$, $(K - 1)$ e $(n - K)$ graus de liberdade.

Isto é: $\chi^2_{n-1} = \chi^2_{K-1} + \chi^2_{n-K}$

Nota-se que: $(n - 1) = (K - 1) + (n - K)$

Sendo esta a condição necessária e suficiente para que χ^2_{K-1} e χ^2_{n-K} sejam independentes.

Dessa maneira,

$$F = \frac{\frac{\chi^2_{K-1}}{K-1}}{\frac{\chi^2_{n-K}}{n-K}} = \frac{\frac{Q_e}{K-1}}{\frac{Q_r}{n-K}} = \frac{S_e^2}{S_r^2}$$

terá distribuição F com $(K - 1)$ $g \cdot \ell$. no numerador e $(n - K)$ $g \cdot \ell$. no denominador.

O quociente F será utilizado para testar a hipótese H_0.

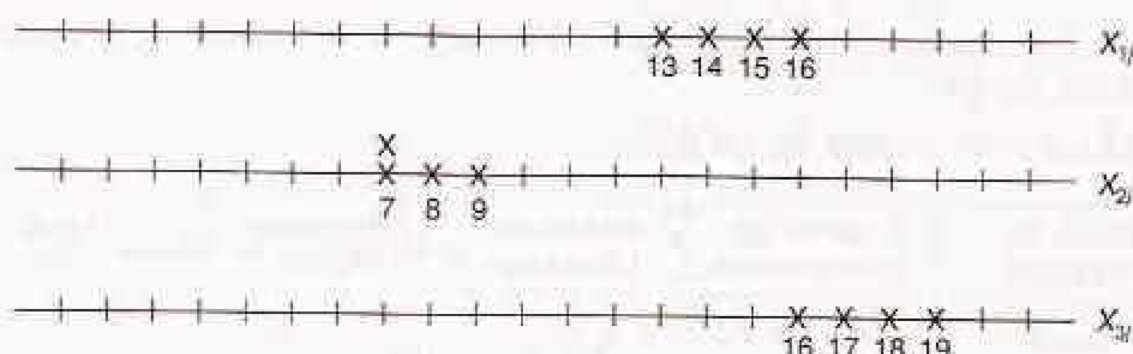
Quanto mais próximo de 1 for o quociente H_0 deverá ser aceita; ao contrário, quanto maior o valor de F, o teste irá indicar a rejeição de H_0, e nesse caso conclui-se com risco α que o fator considerado tem influência sobre a variável dependente.

11.3.2 Fundamentos da Análise da Variância (ANOVA)

A configuração seguinte ilustra a base do método e o conseqüente uso do quociente F para se testar a hipótese da igualdade das médias. Para tanto, vamos supor três amostras de quatro elementos cada uma, cujos valores são:

Amostra 1: χ_{1j} :	14	16	15	13
Amostra 2: χ_{2j} :	7	7	8	9
Amostra 3: χ_{3j} :	16	17	18	19

e que estão mostradas no gráfico:



Vê-se claramente um caso em que a hipótese H_0 será rejeitada pela análise de variância. As três amostras parecem confirmar a hipótese de homocedasticidade (variâncias iguais), todavia, as médias, diferem claramente de amostra para amostra: ($\bar{x}_1$ = 14,5; $\bar{x}_2$ = 7,75; $\bar{x}_3$ = 17,5) calculando-se, dessa forma, os valores das estimativas da variância encontraremos: S_t^2= 19,30; S_e^2 = 99,75; S_r^2 = 1,42.

Nota-se pela análise do gráfico e dos resultados que, sendo H_0 falsa, haverá uma tendência de S_t^2 e S_e^2 superestimarem (19,30 e 99,75, respectivamente) σ^2. O que não ocorre com S_r^2 , já que, S_r^2= 1,42, é uma boa estimativa de σ^2.

Contrariamente, se H_0 for verdadeira, S_t^2, S_e^2 e S_r^2 fornecerão boas estimativas para σ^2. Imagine, olhando para o gráfico com os três grupos alinhados em torno de um eixo vertical ($\bar{x}$).

Tem-se aí o fundamento lógico da análise da variância. Na verdade, o teste de igualdade de médias é substituído por um teste de igualdade de variâncias: $\sigma_e^2 = \sigma_r^2$.

Assim, se a hipótese de igualdade das variâncias for aceita, pode-se concluir que as médias são iguais, pois neste caso o estimador o S_e^2 terá a mesma dimensão que S_r^2, ou seja, o quociente entre ambos estará próximo da unidade. Porém, se a hipótese da igualdade das médias não é verificada se terá S_e^2 bem maior do que S_r^2, e conseqüentemente o valor do quociente será bem maior do que a unidade.

É fácil compreender que o teste da análise da variância será unicaudal à direita, com o risco α concentrado na cauda à direita.

11.3.3 Quadro de Análise da Variância

Os resultados obtidos poderão ser reunidos no Quadro de Análise da Variância, assim:

Quadro de Análise da Variância

Fonte de Variação	***Soma de Quadrados***	***Graus de Liberdade***	***Quadrados Médios***	***Teste F***
Entre Tratamentos	Q_e	$K-1$	$S_e^2 = \frac{Q_e}{K-1}$	$F_{cal} = \frac{S_e^2}{S_r^2}$
Dentro das Amostras (Residual)	$Q_r = Q_t - Q_e$	$n-K$	$S_r^2 = \frac{Q_t - Q_e}{n-K}$	
Total	Q_t	$n-1$		

Para testar a hipótese H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_K = \mu$ contra H_1: $\mu_p \neq \mu_q$, para $p \neq q$, compara-se o valor F_{cal} com o valor F tabelado com $(K-1)$ $g \cdot \ell \cdot$ no numerador e $(n-K)$ no denominador, fixando certo nível α de significância.

Se $F_{cal} \leq F_{tab}$, então aceita-se H_0 e conclui-se com risco α que o fator considerado não causa efeito sobre a variável em estudo. Por outro lado, se $F_{cal} > F_{tab}$, rejeita-se H_0, concluindo-se pela diferença das médias e conseqüente influência do fator sobre a variável analisada.

A seguir apresenta-se um procedimento para a efetivação do teste:

1º) Dispor os elementos, segundo a tabela a seguir, obtendo as somas das colunas e suas respectivas médias.

Elemento da amostra \ Tratamentos	1	2	3	....	K	
1	x_{11}	x_{21}	x_{31}	....	x_{K1}	
2	x_{12}	x_{22}	x_{32}	....	x_{K2}	
.	.	.	.		.	
.	.	.	.		.	
.	.	.	.		.	
n_i	x_{1n_1}	x_{2n_2}	x_{3n_3}	....	x_{Kn_K}	
Σ						
$\bar{x}_i$						$\bar{X}$

2º) Calcula-se a constante

$$C = \frac{\left(\sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij}\right)^2}{n}$$

3º) Avalia-se a soma: $\sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij}^2$, obtendo a variação total

$$Q_t = \sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij}^2 - C$$

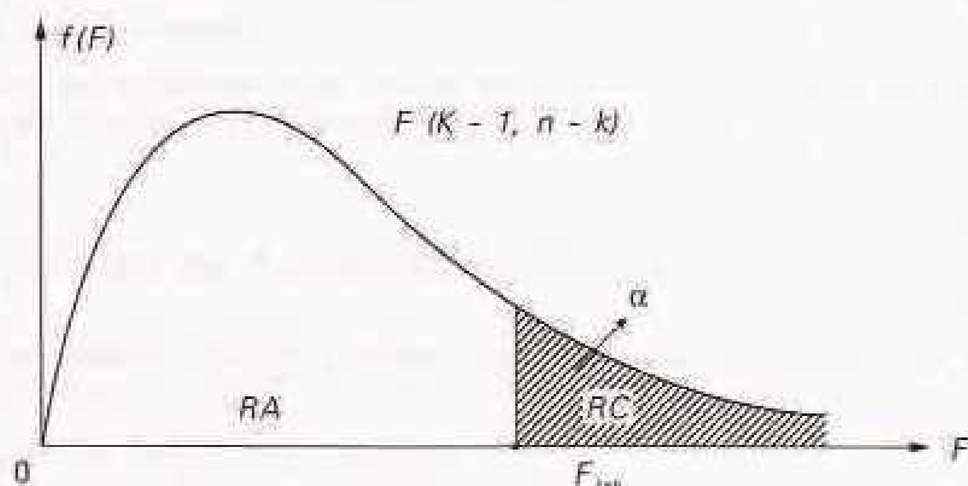
4º) Calcula-se a variação entre tratamentos

$$Q_e = \sum_{i} \left[\frac{\left(\sum_{j} x_{ij} \right)^2}{n_i} \right] - C$$

5º) Obtém-se a variação residual por diferença: $Q_r = Q_t - Q_e$.

6º) Constrói-se o Quadro de Análise da Variância, avaliando o F_{cal}.

7º) Determina-se a região crítica e de aceitação da hipótese H_0 por meio da tabela F.



8º) Compara-se F_{cal} com F_{tab}, obtendo-se a conclusão.

Exemplo: O resultado das vendas efetuadas por 3 vendedores de uma indústria durante certo período é dado a seguir. Deseja-se saber, ao nível de 5%, se há diferença de eficiência entre os vendedores.

Vendedores		
A	***B***	***C***
29	27	30
27	27	30
31	30	31
29	28	27
32		29
30		

Solução: Sem afetar os resultados, pode-se subtrair uma constante, digamos 28, a todos os valores, simplificando dessa maneira os cálculos.

Assim:

Vendedores			
A	*B*	*C*	
1	–1	2	
–1	–1	2	
3	2	3	
1	0	–1	
4		1	
2			Total
Σ 10	0	7	17

2º) $C = \frac{(17)^2}{15} = 19,27$

3º) $\sum_{i=1}^{3} \sum_{j=1}^{n_i} x_{ij}^2 = 1^2 + (-1)^2 + 3^2 + 1^2 + \ldots + (-1)^2 + 1^2 = 57$

$Q_t = 57 - 19,27 = 37,73$

4º) $Q_e = \sum_{i=1}^{3} \left[\frac{\left(\sum_{j=1}^{n_i} x_{ij} \right)^2}{n_i} \right] - C$

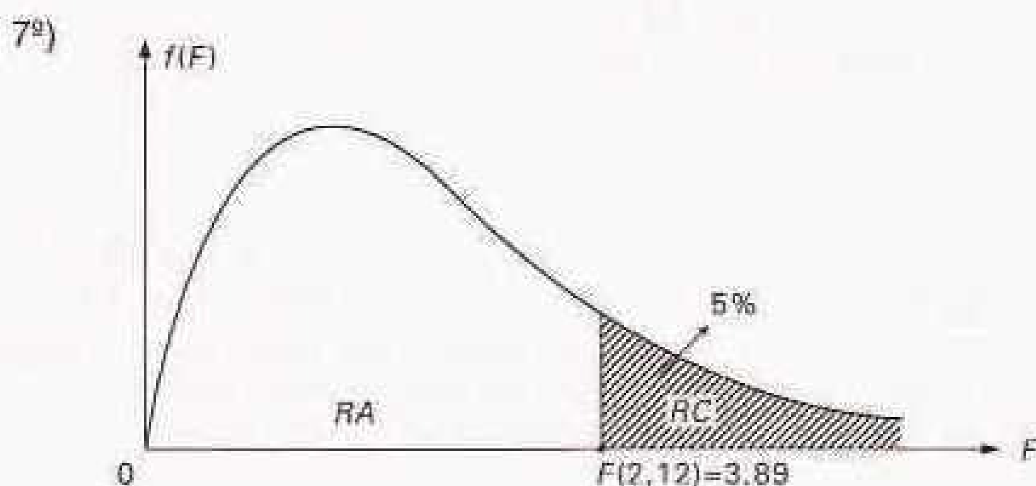
$= \frac{(10)^2}{6} + \frac{(0)^2}{4} + \frac{(7)^2}{5} - 19,27 = 7,20$

5º) $Q_r = Q_t - Q_e$; $Q_r = 37,73 - 7,20 = 30,53$

6º) QAV

Fonte de Variação	Soma de Quadrados	G.L.	Quadrados Médios	Teste F
Entre os Tratamentos	$Q_e = 7{,}20$	$K-1$ $3-1=2$	$S_e^2 = \frac{7{,}2}{2} = 3{,}6$	$F_{cal} = \frac{3{,}6}{2{,}54} = 1{,}41$
Residual	$Q_r = 30{,}53$	$n-K$ $15-3=12$	$S_r^2 = \frac{30{,}53}{12} = 2{,}54$	
Total	$Q_t = 37{,}73$	$n-1$ $15-1=14$		

7º)



8º) Como $F_{cal} = 1{,}42 < F_{tab} = 3{,}89$, aceita-se H_0, concluindo-se com nível de 5% que não há diferença entre os vendedores. Isto é, aceita-se H_0: $\mu_{(A)} = \mu_{(B)} = \mu_{(C)}$.

11.4 CLASSIFICAÇÃO DE DOIS CRITÉRIOS OU EXPERIMENTO DE DOIS FATORES

Admitem-se dois fatores (variáveis independentes). A variável de estudo (variável dependente) é observada em cada cela, combinação dos tratamentos do fator 1, e dos blocos do fator 2. Tem-se uma tabela de "K" colunas e "L" linhas. Ou seja, $K \cdot L = n$ observações.

	Primeiro critério (colunas) = tratamentos			
Segundo critério (linhas) Blocos	x_{11}	x_{21}	...	x_{K1}
	x_{12}	x_{22}	...	x_{K2}
	.	.		.
	.	.		.
	.	.		.
	x_{1L}	x_{2L}	...	x_{KL}

Considere um experimento de natureza agrícola consistindo no exame das safras por are de 3 variedades de soja, em que cada variedade é plantada em 4 lotes diferentes de terra. Tem-se um total de (3) (4) = 12 lotes. Em tal caso é conveniente combinar os lotes em blocos, digamos 3 lotes constituindo um bloco, com uma variedade distinta de soja plantada em cada lote do bloco. São, então, necessários 4 blocos. Neste caso, há duas classificações ou fatores, pois pode haver diferença na produção por are devida: ao tipo de soja; ou ao particular bloco considerado.

Denota-se por $\bar{x}_i$ a média de uma coluna i qualquer, por $\bar{x}_j$ a média de uma linha j qualquer e por $\bar{X}$ a média geral:

$$\bar{x}_i = \frac{1}{L} \sum_{j=1}^{L} x_{ij} \qquad \bar{x}_j = \frac{1}{K} \sum_{i=1}^{K} x_{ij} \qquad \bar{X} = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{L} x_{ij}$$

Assim, como no caso da classificação única, admite-se que todas as amostras provenham de populações normais com a mesma variância:

Para a comparação das médias entre colunas (tratamentos) a hipótese nula será H_0^C: $\mu_i = \mu$ para qualquer $i = 1, 2, ..., K$ será testada contra a hipótese alternativa H_1^C: $\mu_i \neq \mu$.

Analogamente, para a comparação das médias entre linhas (blocos), a hipótese que será colocada à prova será H_0^L: $\mu_j = \mu$ para qualquer $j = 1, 2, ..., L$, enquanto H_1^L : $\mu_j \neq \mu$.

11.4.1 Estimadores da Variância Comum σ^2

Pode-se estimar a variância de σ^2 de 4 formas diferentes. A estimativa total S_t^2, a estimativa entre linhas, S_L^2, a estimativa entre colunas S_C^2, e a estimativa residual, S_r^2. Assim:

1º) Estimativa total: S_t^2

$$S_t^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{L} (x_{ij} - \bar{x})^2}{n - 1} = \frac{Q_t}{n - 1}$$

O numerador (Q_t) representa a variação total. Por outro lado, sabe-se que $\frac{Q_t}{\sigma^2}$ tem distribuição χ^2 com $(n - 1)$ $g \cdot \ell$. A fórmula prática da variação total é dada por:

$$Q_t = \sum_{i=1}^{K} \sum_{j=1}^{L} x_{ij}^2 - C$$

$$C = \frac{\left(\sum_i \sum_j x_{ij}\right)^2}{n}$$

2º) Estimativa entre colunas: S_C^2

$$S_C^2 = L \frac{\sum_{i=1}^{K} (\bar{x}_i - \bar{x})^2}{K - 1} = \frac{Q_{EC}}{K - 1}$$

O numerador (Q_{EC}) representa a variação entre colunas. Demonstra-se que $\frac{Q_{EC}}{\sigma^2}$ tem distribuição χ^2 com $(K - 1)$ $g \cdot \ell$. A fórmula prática para o cálculo de Q_{EC} é dada por:

$$Q_{EC} = \sum_{i=1}^{K} \left[\frac{\left(\sum_j x_{ij}\right)^2}{L} \right] - C$$

3º) **Estimativa entre linhas:** S_L^2

$$S_L^2 = K \frac{\sum_{j=1}^{L} (\bar{x}_j - \bar{\bar{x}})^2}{L-1} = \frac{Q_{EL}}{L-1}$$

A variação entre colunas é dada pelo numerador da expressão. Da mesma maneira, $\frac{Q_{EL}}{\sigma^2}$ tem distribuição χ^2 com $(L-1)$ $g \cdot \ell$. A fórmula prática para seu cálculo é dada por:

$$\boxed{Q_{EL} = \sum_{j=1}^{L} \left[\frac{\left(\sum_{i} x_{ij} \right)^2}{K} \right] - C}$$

4º) **Estimativa residual:** S_r^2

$$S_r^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{L} [(x_{ij} - \bar{\bar{x}}) - (\bar{x}_i - \bar{\bar{x}}) - (\bar{x}_j - \bar{\bar{x}})]^2}{(K-1)(L-1)}$$

ou

$$S_r^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{L} (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2}{(K-1)(L-1)} = \frac{Q_r}{(K-1)(L-1)}$$

A variação residual é dada pelo numerador e também neste caso $\frac{Q_r}{\sigma^2}$ tem distribuição χ^2 com $(K-1)$ $(L-1)$ graus de liberdade. A avaliação de Q_r é obtida por diferença, já que também neste caso é válida a igualdade:

$$\boxed{Q_t = Q_{EC} + Q_{EL} + Q_r}$$

Assim:

$$\boxed{Q_r = Q_t - Q_{EC} - Q_{EL}}$$

Convém observar também neste caso que S_C^2; S_L^2 serão estimadores justos se tanto H_0^C como H_0^L forem verdadeiras, ao passo que S_r^2 será um estimador justo sob quaisquer hipóteses sobre o comportamento das médias.

Por outro lado, nota-se que, se

$$Q_t = Q_{EC} + Q_{EL} + Q_r$$

então

$$\sigma^2\chi^2_{n-1} = \sigma^2\chi^2_{K-1} + \sigma^2\chi^2_{L-1} + \sigma^2\chi^2_{(K-1)(L-1)}$$

ou, ainda,

$$\chi^2_{n-1} = \chi^2_{K-1} + \chi^2_{L-1} + \chi^2_{(K-1)(L-1)}$$

Nota-se que a soma dos graus de liberdade dos qui-quadrados do segundo membro é igual ao número de $g \cdot \ell$ do qui-quadrado particionado, isto é,

$$n - 1 = (K - 1) + (L - 1) + (K - 1)(L - 1)$$
$$n - 1 = K - 1 + L - 1 + KL - K - L + 1$$
$$n - 1 = KL - 1$$

Lembre-se de que $KL = n$. Logo:

$$\chi^2_{K-1};\ \chi^2_{L-1} \text{ e } \chi^2_{(K-1)(L-1)}$$

são qui-quadrados independentes, e, dessa maneira, pode-se testar a igualdade das médias segundo as colunas e/ou linhas mediante o cálculo de

para colunas:

$$\boxed{F^C_{\text{cal}} = \frac{S_C^2}{S_r^2};}$$

para linhas:

$$\boxed{F^L_{\text{cal}} = \frac{S_L^2}{S_r^2}.}$$

Deve-se, também, salientar que o fato de H_0^C não ser verdadeira não impede que se teste H_0^L, e vice-versa. O quadro da análise da variância a seguir resume ambos os testes.

QAV

Fonte de Variação	*Soma dos Quadrados*	*G.L.*	*Quadrados Médios*	*Teste F*
Entre colunas	$Q_{EC}=\sum_i\left[\frac{\left(\sum_j x_{ij}\right)^2}{L}\right]-C$	$K-1$	$S_C^2=\frac{Q_{EC}}{K-1}$	
Entre linhas	$Q_{EL}=\sum_i\left[\frac{\left(\sum_j x_{ij}\right)^2}{K}\right]-C$	$L-1$	$S_L^2=\frac{Q_{EL}}{L-1}$	$F_{cal}^C=\frac{S_C^2}{S_r^2}$
Residual	$Q_r=Q_t-Q_{EC}-Q_{EL}$	$(K-1)(L-1)$	$S_r^2=\frac{Q_r}{(K-1)(L-1)}$	$F_{cal}^L=\frac{S_L^2}{S_r^2}$
Total	$Q_t=\sum_i\sum_j x_{ij}^2-C$	$n-1$		

Regra de decisão: Fixando certo nível de significância α, tem-se:

1. Se $F_{cal}^C \leq F[(K-1);(K-1)(L-1)]$, então, aceita-se H_{0_C}: $\mu_i=\mu$ para qualquer $i=1, 2, ..., K$, e conclui-se com risco α que o fator 1 (tratamentos) não causa efeito na variável dependente. Por outro lado, se $F_{cal}^C > F_{tab}$, rejeita-se H_0, concluindo-se pela diferença das médias das colunas e conseqüente influência do fator sobre a variável analisada.
2. Se $F_{cal}^L \leq F[(L-1);(K-1)(L-1)]$, então, aceita-se H_0^L: $\mu_j=\mu$ para qualquer $j=1, 2, ..., L$, e conclui-se com risco α que o fator 2 (blocos) não causa efeito na variável dependente. Por outro lado, se $F_{cal}^L > F_{tab}$, rejeita-se H_0, concluindo-se pela diferença das médias das linhas e conseqüente influência do fator sobre a variável em estudo.

Encontram-se, a seguir, os principais passos para a efetivação do teste:

1. Dispor os elementos segundo a tabela que segue. Obtendo as somas das colunas e linhas, bem como suas respectivas médias:

Fator 1 (i) / Fator 2 (j)	1	2	....	K	$\sum$	$\bar{x}_j$
1	x_{11}	x_{21}	....	x_{K1}		
2	x_{12}	x_{22}	....	x_{K2}		
.	.	.	.	.		
.	.	.	.	.		
.	.	.	.	.		
L	x_{1L}	x_{2L}	....	x_{KL}		
Σ						
$\bar{x}_i$						

2. Avalia-se a constante

$$C = \frac{\left(\sum_{i=1}^{K} \sum_{j=1}^{L} x_{ij} \right)^2}{n}$$

Lembre-se de que $n = KL$

3. Calcula-se a variação total

$$Q_t = \sum_{i=1}^{K} \sum_{j=1}^{L} x_{ij}^2 - C$$

4. Determina-se a variação entre colunas

$$Q_{EC} = \sum_{i} \left[\frac{\left(\sum_{j} x_{ij} \right)^2}{L} \right] - C$$

5. Avalia-se a variação entre linhas

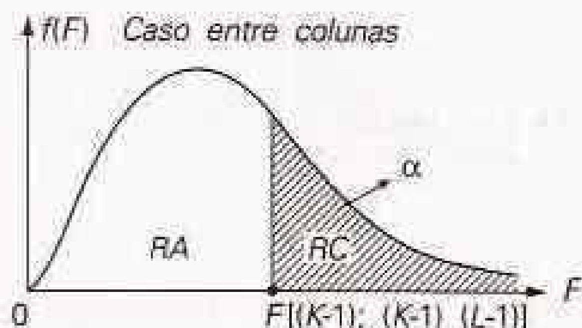
$$Q_{EL} = \sum_j \left[\frac{\left(\sum_i x_{ij} \right)^2}{K} \right] - C$$

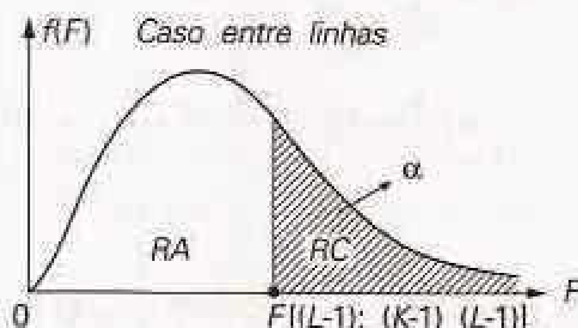
6. Obtém-se a variação residual por diferença

$$Q_r = Q_t - Q_{EC} - Q_{EL}$$

7. Constrói-se o quadro de análise da variância, avaliando-se F^C_{cal} e F^L_{cal}.

8. Determina-se a RC e a RA por meio da tabela F.





9. Efetuam-se as conclusões pela comparação dos respectivos valores dos *F* calculados e tabelados.

Exemplo: Em uma experiência agrícola, foram usados cinco diferentes fertilizantes em duas variedades de trigo. A produção está indicada a seguir. Verificar ao nível de 5% se: a) há diferença na produção devido ao fertilizante; b) há diferença na safra devido à variedade do trigo.

Fertilizante	A	B	C	D	E
Variedade 1	54	38	46	50	44
Variedade 2	57	42	45	53	50

Solução: Considerando-se o fator 1 como o tipo de fertilizante e o fator 2 como variedade de trigo, constrói-se a tabela, subtraindo 45 a todos os valores observados. Assim:

(i) Fator 1 / *(j) Fator 2*	*A*	*B*	*C*	*D*	*E*	Σ
1	9	–7	1	5	–1	7
2	12	–3	0	8	5	22
Σ	21	–10	1	13	4	29

2. $$C = \frac{\left(\sum_{K=1}^{5}\sum_{L=1}^{2} x_{ij}\right)^2}{10} = \frac{(29)^2}{10} = 84{,}1$$

3. $$Q_t = \sum_{K=1}^{5}\sum_{L=1}^{2} x_{ij}^2 - C$$

$$= 9^2 + 12^2 + (-7)^2 + \ldots + 5^2 - 84{,}1 = 314{,}9$$

4. $$Q_{EC} = \sum_{i=1}^{5}\left[\frac{\left(\sum_{j=1}^{2} x_{ij}\right)^2}{2}\right] - C$$

$$= \frac{(21)^2}{2} + \frac{(-10)^2}{2} + \frac{(1)^2}{2} + \frac{(13)^2}{2} + \frac{(4)^2}{2} - 84{,}1 = 279{,}4$$

5. $$Q_{EL} = \sum_{j=1}^{2}\left[\frac{\left(\sum_{i=1}^{5} x_{ij}\right)^2}{5}\right] - C$$

$$= \frac{(7)^2}{5} + \frac{(22)^2}{5} - 84{,}1 = 22{,}5$$

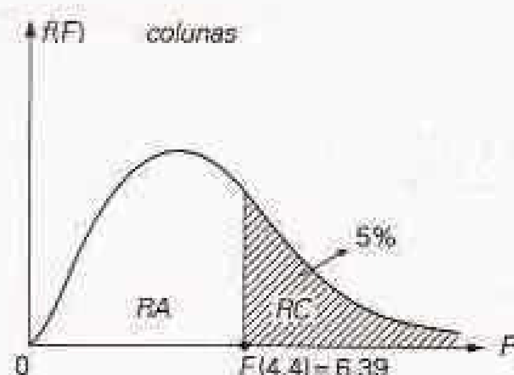
6. $Q_r = Q_t - Q_{EC} - Q_{EL}$

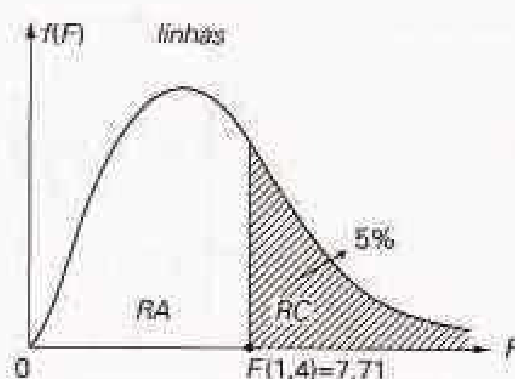
$Q_r = 314{,}9 - 279{,}4 - 22{,}5 = 13$

7. QAV

Fonte de variação	*Soma dos quadrados*	*G.L.*	*Quadrados médios*	*Teste F*
Entre colunas	279,4	4	69,85	$F_{cal}^{C} = \frac{69,85}{3,25} = 21,49$
Entre linhas	22,5	1	22,5	$F_{cal}^{L} = \frac{22,5}{3,25} = 6,92$
Residual	13	4	3,25	
Total	314,9	9		

8. RC e RA





9. *Conclusão:* Para o primeiro fator, fertilizante, tem-se que $F_{cal}^{C} = 21,49 \geq 6,39$; portanto, rejeita-se H_0: $\mu_i = \mu$; $i = 1,2,...,5$, concluindo-se que o tipo de fertilizante tem influência na produção de trigo.

Para o segundo fator, variedade de trigo, tem-se que $F_{cal}^{L} = 6,92 < 7,71$; portanto, aceita-se H_0: $\mu_j = \mu$; $j = 1, 2$, concluindo-se que a variedade de trigo não altera a produção.

11.5 EXPERIMENTO DE DOIS FATORES COM REPETIÇÃO

Nesse caso haverá mais de um valor correspondendo a um tratamento e um bloco. Admite-se que haja R valores para cada posição. Tem-se K colunas (tratamentos); L linhas (blocos) e R observações para cada interação. Assim, $n = KLR$, sendo $x_{ij\ell}$ o elemento correspondente à coluna de ordem i, à linha j, e à repetição de ordem ℓ, sendo $i = 1, 2, ..., K$; $j = 1, 2, ..., L$; $\ell = 1, 2, ..., R$.

Como anteriormente, a variação total Q_t pode decompor-se em variação devido às colunas: Q_{EC}; variação devido às linhas: Q_{EL}; variação devido à interação: Q_I e variação residual: Q_r; assim:

$$Q_t = Q_{EC} + Q_{EL} + Q_I + Q_r$$

onde:

a) variação total

$$Q_t = \sum_{i=1}^{K} \sum_{j=1}^{L} \sum_{\ell=1}^{R} (x_{ij\ell} - \bar{x})^2 \quad \text{ou} \quad Q_t = \sum_{i=1}^{K} \sum_{j=1}^{L} \sum_{\ell=1}^{R} x_{ij\ell}^2 - C$$

em que

$$C = \frac{\left(\sum_i \sum_j \sum_\ell x_{ij\ell}\right)^2}{n}$$

b) variação entre tratamento

$$Q_{EC} = \sum_i \sum_j \sum_\ell (\bar{x}_i - \bar{x})^2 = RL \sum_i (\bar{x}_i - \bar{x})^2$$

ou

$$Q_{EC} = \sum_{i=1}^{K} \left[\frac{\left(\sum_{j=1}^{L} \sum_{\ell=1}^{R} x_{ij\ell} \right)^2}{RL} \right] - C$$

c) variação entre blocos

$$Q_{EL} = \sum_{i} \sum_{j} \sum_{\ell} (\bar{x}_j - \bar{x})^2 = RK \sum_{j=1}^{L} (\bar{x}_j - \bar{x})^2$$

ou

$$Q_{EL} = \sum_{j=1}^{L} \left[\frac{\left(\sum_{i} \sum_{\ell} x_{ij\ell} \right)^2}{RK} \right] - C$$

d) variação residual

$$Q_r = \sum_{i} \sum_{j} \sum_{\ell} (x_{ij\ell} - \bar{x}_{ij})^2$$

ou

$$Q_r = \sum_{i=1}^{K} \sum_{j=1}^{L} \sum_{\ell=1}^{R} x_{ij\ell}^2 - \sum_{i=1}^{K} \sum_{j=1}^{L} \left[\frac{\left(\sum_{\ell=1}^{R} x_{ij\ell} \right)^2}{R} \right]$$

e) variação devido a interação

$$Q_I = \sum_{i} \sum_{j} \sum_{K} (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$$

ou

$$Q_I = R \sum_{i} \sum_{j} (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$$

Praticamente, seu valor será obtido por diferença:

$$\boxed{Q_I = Q_t - Q_{EC} - Q_{EL} - Q_r}$$

Usando o número apropriado de graus de liberdade para cada fonte de variação, pode-se montar o quadro de análise da variância.

QAV

Fonte de variação	*Soma dos quadrados*	*G.L.*	*Quadrados médios*	*Teste F*
Entre colunas	Q_{EC}	$K-1$	$S_C^2 = \dfrac{Q_{EC}}{K-1}$	$F_{cal}^C = \dfrac{S_C^2}{S_r^2}$
Entre linhas	Q_{EL}	$L-1$	$S_L^2 = \dfrac{Q_{EL}}{L-1}$	
Devido à interação	Q_I	$(L-1)(K-1)$	$S_I^2 = \dfrac{Q_I}{(K-1)(L-1)}$	$F_{cal}^L = \dfrac{S_L^2}{S_r^2}$
Residual	Q_r	$LK(R-1)$	$S_r^2 = \dfrac{Q_r}{LK(R-1)}$	$F_{cal}^I = \dfrac{S_I^2}{S_r^2}$
Total	Q_t	$n-1 = KLR-1$	$S_t^2 = \dfrac{Q_t}{n-1}$	

Os F_{cal} da última coluna podem ser usados para testar as hipóteses nulas:

H_0^C: Todas as médias de tratamento (colunas) são iguais;

H_0^L: Todas as médias de blocos (linhas) são iguais;

H_0^I: Não há interações entre tratamentos e blocos.

Eis o procedimento para se testarem essas três hipóteses:

1. Dispor os dados da maneira exposta na tabela seguinte. Obter as somas das celas, das linhas e das colunas.

Blocos \ Tratamentos	1	2	···	K	Σ
1	x_{111} x_{112} ··· x_{11R}	x_{211} x_{212} ··· x_{21R}	···	$x_{K_{11}}$ ··· x_{K_1R}	
2	x_{121} x_{122} ··· x_{12R}	x_{221} x_{222} ··· x_{22R}	···	$x_{K_{21}}$ ··· x_{K_2R}	
.	.	.		.	
.	.	.		.	
.	.	.		.	
L	x_{1L1} x_{1L2} ··· x_{1LR}	x_{2L1} x_{2L2} ··· x_{2LR}	···	x_{KL1} ··· x_{KLR}	
Σ					

2. Avaliar a constante

$$C = \frac{\left(\sum_i \sum_j \sum_\ell x_{ij\ell} \right)^2}{n}$$

bem como a variação total

$$Q_t = \sum_i \sum_j \sum_\ell x_{ij\ell}^2 - C$$

3. Calcular a variação entre tratamentos (colunas):

$$Q_{EC} = \sum_{i=1}^{K} \left[\frac{\left(\sum_i \sum_\ell x_{ij\ell} \right)^2}{RK} \right] - C$$

4. Determinar a variação entre blocos (linhas):

$$Q_{EL} = \sum_{j=1}^{L} \left[\frac{\left(\sum_i \sum_\ell x_{ij\ell} \right)^2}{RK} \right] - C$$

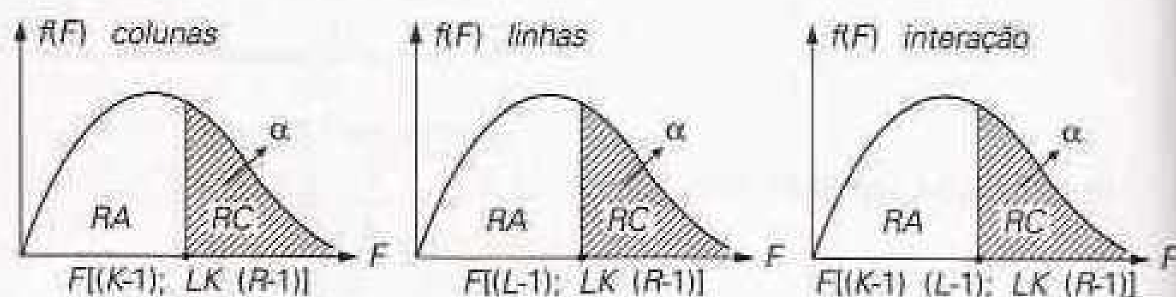
5. Determinar a variação residual:

$$Q_r = \sum_i \sum_j \sum_\ell x_{ij\ell}^2 - \sum_i \sum_j \left[\frac{\left(\sum_\ell x_{ij\ell} \right)^2}{R} \right]$$

6. Avaliar a variação devido à interação

$$Q_I = Q_t - Q_{EC} - Q_{EL} - Q_r$$

7. Fixado um nível α de significância, determinar as regiões crítica e de aceitação para cada teste.



8. Construir o Q.A.V.
9. Fazer a conclusão, comparando os respectivos F_{cal} com as regiões crítica e de aceitação.

Exemplo: As compras de chá-mate de 18 famílias estão dadas a seguir. Cada família está classificada segundo a cidade em que reside e o número de vezes que foi exposta à propaganda do chá noticiada pela TV. Para se conhecer a evolução do efeito da propaganda, deseja-se saber ao nível de 5%:

a) se há alguma relação entre propaganda noticiada e consumo do produto;

b) se há alguma diferença significativa para o consumo entre as cidades;

c) se há alguma relação entre propaganda noticiada e as cidades.

Cidades	Número de vezes de colocação da propaganda					
	de 1 a 5 vezes		de 6 a 10 vezes		mais de 10 vezes	
A	19	27	18	20	30	18
B	18	26	27	19	25	32
C	24	21	19	31	25	30

Solução:

1. Subtraindo-se a todos os valores a constante 25, obteremos a tabela:

Fator 1 / Fator 2			Σ			Σ			Σ	Σ
	−6	2	−4	−7	−5	−12	5	−7	−2	−18
	−7	1	−6	2	−6	−4	0	7	7	−3
	−1	−4	−5	−6	6	0	0	5	5	0
Σ			−15			−16			10	−21

2. $$C = \frac{\left(\sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{\ell=1}^{2} x_{ij\ell} \right)^2}{n = KLR} = \frac{(-21)^2}{18} = 24{,}5$$

$$Q_t = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{\ell=1}^{2} x_{ij\ell}^2 - C$$

$$= (-6)^2 + (2)^2 + (-7)^2 + (1)^2 + \dots + (0)^2 + (5)^2 - 24{,}5 = 416{,}5$$

3. $$Q_{EC} = \sum_{i=1}^{3} \left[\frac{\left(\sum_{j=1}^{3} \sum_{\ell=1}^{2} x_{ij\ell} \right)^2}{RL} \right] - C$$

$$= \frac{(-15)^2}{6} + \frac{(-16)^2}{6} + \frac{(10)^2}{6} - 24{,}5 = 72{,}33$$

4. $Q_{EL} = \sum_{j=1}^{3} \left[\frac{\left(\sum_{i=1}^{3} \sum_{\ell=1}^{2} x_{ij\ell} \right)^2}{RK} \right] - C$

$$= \frac{(-18)^2}{6} + \frac{(-3)^2}{6} + \frac{(0)^2}{6} - 24{,}5 = 31$$

5. $Q_r = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{\ell=1}^{3} x_{ij\ell}^2 - \sum_{i=1}^{3} \sum_{j=1}^{3} \left[\frac{\left(\sum_{\ell=1}^{2} x_{ij\ell} \right)^2}{R} \right] =$

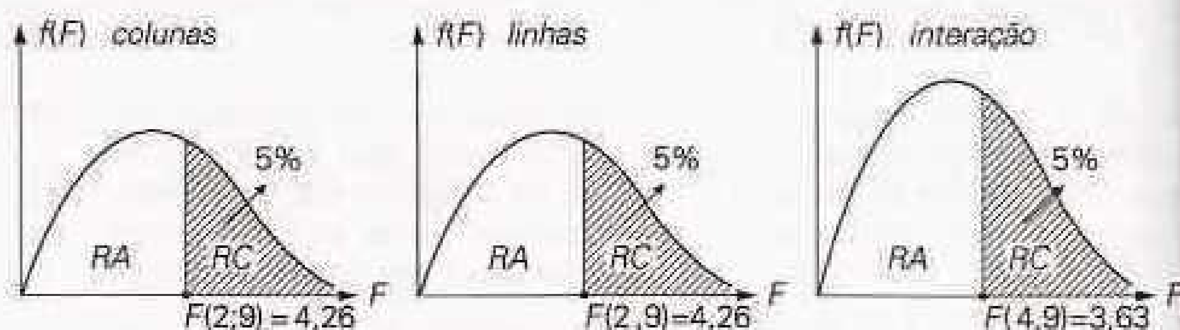
$$= (-6)^2 + (2)^2 + (-7)^2 + \dots + (5)^2 - \left[\frac{(-4)^2}{2} + \frac{(-6)^2}{2} + \right.$$

$$\left. + \frac{(-5)^2}{2} + \frac{(-12)^2}{2} + \dots + \frac{(5)^2}{2} \right] = 283{,}5$$

6. $Q_I = Q_t - Q_{EC} - Q_{EL} - Q_r$

$Q_I = 416{,}5 - 72{,}33 - 31 - 283{,}5 = 29{,}67$

7. Com o auxílio da tabela F, determinam-se as regiões crítica e de aceitação.



8. **QAV**

Fonte de variação	*Soma dos quadrados*	*G.L.*	*Quadrados médios*	*Teste F*
Entre colunas	72,33	2	36,17	$F_{cal}^{C} = \frac{36{,}17}{31{,}5} = 1{,}15$
Entre linhas	31	2	15,5	$F_{cal}^{L} = \frac{15{,}5}{31{,}5} = 0{,}49$
Devido à interação	29,67	4	7,42	$F_{cal}^{I} = \frac{7{,}42}{31{,}5} = 0{,}24$
Residual	283,5	9	31,5	
Total	416,5	17		

9. **Conclusão**:

a) Como $F_{cal}^{C} < F_{tab}$, aceita-se H_0, concluindo-se com risco de 5% que não há relação entre propaganda noticiada e consumo de chá.

b) Como $F_{cal}^{L} < F_{tab}$, aceita-se H_0, concluindo-se com risco de 5% que não há diferença para o consumo entre cidades.

c) Como $F_{cal}^{I} < F_{tab}$, aceita-se H_0, concluindo-se com risco de 5% que não há relação entre propaganda e cidades.

11.6 TESTE DE SCHEFFÉ

O método de análise de variância indica a aceitação ou rejeição da hipótese de igualdade das médias. Se H_0 for rejeitada, pelo menos uma das médias é diferente das demais. Surge, contudo, a questão: Quais médias devem ser consideradas diferentes?

Existem alguns testes para a solução dessa questão; apresentaremos um deles: o teste de Scheffé.

Segundo esse procedimento, devem ser consideradas distintas entre si, ao nível de significância adotado, as médias μ_A e μ_B, tais que:

1. Para o caso do modelo de classificação única:

$$|\bar{X}_A - \bar{X}_B| > \sqrt{S_r^2(K-1)\left(\frac{1}{n_A} + \frac{1}{n_B}\right) F\alpha\,[(K-1);(n-K)]}$$

2. Para o caso de duas classificações:

para as colunas:

$$|\bar{X}_A - \bar{X}_B| > \sqrt{S_r^2 \frac{2(K-1)}{L} F_\alpha[(K-1); (K-1)(L-1)]}$$

para as linhas:

$$|\bar{X}_A - \bar{X}_B| > \sqrt{S_r^2 \frac{2(L-1)}{K} F_\alpha[(L-1); (K-1)(L-1)]}$$

3. Para o caso de duas classificações com repetição:

para as colunas:

$$|\bar{X}_A - \bar{X}_B| > \sqrt{S_r^2 \frac{2(K-1)}{RL} F_\alpha[(K-1); KL(R-1)]}$$

para as linhas:

$$|\bar{X}_A - \bar{X}_B| > \sqrt{S_r^2 \frac{2(L-1)}{RK} F_\alpha[(L-1); KL(R-1)]}$$

Exemplo: Vamos descobrir os pares de médias diferentes para o exemplo de classificação dupla visto anteriormente. Naquele caso, o primeiro fator – fertilizante – tem influência na produção de trigo.

Relembrando:

Fertilizante	***A***	***B***	***C***	***D***	***E***
Variedade 1	54	38	46	50	44
Variedade 2	57	42	45	53	50
Σ	111	80	91	103	94
Médias das colunas x_j	55,5	40	45,5	51,5	47

$|\bar{X}_A - \bar{X}_B| = |55{,}5 - 40| = 15{,}5$

$|\bar{X}_A - \bar{X}_C| = |55{,}5 - 45{,}5| = 10$

$|\bar{X}_A - \bar{X}_D| = |55{,}5 - 51{,}5| = 4$

$|\bar{X}_A - \bar{X}_E| = |55{,}5 - 47| = 8{,}5$

$|\bar{X}_B - \bar{X}_C| = |40 - 45{,}5| = 5{,}5$

$|\bar{X}_B - \bar{X}_D| = |40 - 51{,}5| = 11{,}5$

$|\bar{X}_B - \bar{X}_E| = |40 - 47| = 7$

$|\bar{X}_C - \bar{X}_D| = |45{,}5 - 51{,}5| = 6$

$|\bar{X}_C - \bar{X}_E| = |45{,}5 - 47| = 1{,}5$

$|\bar{X}_D - \bar{X}_E| = |51{,}5 - 47| = 4{,}5$

Então:

$$\Delta = \sqrt{S_r^2 \frac{2(K-1)}{L} F_\alpha[(K-1);(K-1)(L-1)]}$$

$$\Delta = \sqrt{3{,}25 \frac{(2)(4)}{2} F_{5\%}[(5-1);(5-1)(2-1)]} = \sqrt{3{,}25\,(4)\,6{,}39}$$

$$\Delta = 9{,}11$$

comparando-se as diferenças das médias com o valor de $\Delta = 9{,}11$, conclui-se que os pares de médias(μ_A; μ_B); (μ_A; μ_C); (μ_B; μ_D) são diferentes.

11.7 PLANEJAMENTO EXPERIMENTAL

Experimentos podem ser definidos como: "estudos cuja implementação envolve a intervenção do investigador além daquela necessária à sua medição".

O planejamento experimental consiste na escolha dos fatores, determinação do grupo de controle, escolha do modelo a ser usado, e, fundamentalmente, o planejamento dos cuidados e detalhes que devem ser observados para minimizar os efeitos de outras fontes de variação (outras variáveis) que não estão sendo consideradas no estudo.

A Análise da Variância é, particularmente, utilíssima para o tratamento de dados oriundos de planejamentos experimentais. É também utilizada nos planejamentos de estudos observacionais, onde o investigador apenas observa e mede o fenômeno estudado.

EXERCÍCIOS – SÉRIE I – CAPÍTULO 11

1. Quatro analistas determinaram o rendimento de dado processo, obtendo:

Analistas			
1	*2*	*3*	*4*
26	17	36	20
27	20	33	18
24	22	31	17
25	21	29	22
29			21
28			23

Determine: a) as médias para os diferentes analistas;

b) a média total;

c) a variação total;

d) a variação entre tratamentos;

e) a variação dentro dos tratamentos (residual);

f) se há diferença entre as médias, adote $\alpha = 5\%$;

g) se possível, identifique os pares de médias diferentes, usando o teste de Scheffé.

2. Uma companhia deseja testar quatro tipos diferentes de válvulas *A, B, C, D*. As vidas médias em horas constam da tabela que segue, em que cada tipo foi testado aleatoriamente em seis aparelhos idênticos. Teste se há diferença significativa entre as válvulas, ao nível de 5%.

A	53	58	56	60	51	55
B	52	60	52	58	50	54
C	51	57	55	53	54	50
D	49	54	52	50	53	51

3. São feitas cinco misturas da mesma liga metálica e para cada mistura serão efetuadas seis determinações da densidade. Os resultados são:

	Densidades					
Mistura A	3,6	3,5	3,7	3,1	3,1	3,2
Mistura B	3,3	3,5	3,4	3,2	3,4	3,4
Mistura C	3,5	3,3	3,4	3,4	3,3	3,2
Mistura D	3,5	3,4	3,0	3,3	3,3	3,8
Mistura E	3,7	3,4	3,6	3,5	3,6	3,4

Há evidência de que certas misturas tenham densidade média maior do que outra? $\alpha = 5\%$.

4. Os dados a seguir representam, em segundos, o tempo gasto por cinco operários para realizar certa tarefa, usando três máquinas diferentes. Considerando $\alpha = 5\%$, verifique se há diferenças entre as máquinas e entre os operários.

Operários	*Máquinas*		
	A	*B*	*C*
1	40	59	42
2	39	55	51
3	47	55	45
4	45	50	40
5	52	52	41

5. Plantam-se quatro tipos diferentes de sementes de café em cinco blocos. Cada bloco é dividido em quatro lotes, pelos quais se distribuem, então, aleatoriamente, os quatro tipos de sementes. Ao nível de significância de 0,05, teste se a produção, indicada na tabela, varia significativamente:

a) devido ao solo (isto é, os cinco blocos);

b) devido à variedade do café.

		Tipos de café			
		I	*II*	*III*	*IV*
	A	15	12	10	14
	B	19	15	12	11
Blocos	C	18	14	15	12
	D	16	11	12	16
	E	17	16	11	14

6. A seguir estão anotadas as quantidade vendidas de certo artigo, considerando-se três preços de venda e três tipos de distribuidores:

Preços / Distribuidores	$P_1 = 54$	$P_2 = 49$	$P_3 = 44$
Farmácias	78 – 76 74 – 77	108 – 106 110 – 104	124 – 122 123 – 125
Drogarias	78 – 78 80 – 77	108 – 110 111 – 107	126 – 125 122 – 128
Outros	80 – 78 79 – 81	110 – 106 108 – 111	128 – 130 126 – 129

a) Testar se a distribuição interfere nas quantidades vendidas;

b) testar se o preço interfere nas quantidades vendidas;

c) testar o efeito da interação.

Observação: Adote $\alpha = 5\%$ para os três testes.

7. Três técnicos fazem três determinações de dureza em cada um de quatro blocos de certo metal. Ao nível de 5% determine se

a) as durezas médias dos blocos são constantes;

b) as determinações dos técnicos são iguais;

c) há alguma interação entre técnicos e blocos.

Técnicos	Blocos											
	1			*2*			*3*			*4*		
X	516	513	514	517	513	513	512	508	511	506	505	506
Y	529	517	519	513	510	511	509	512	513	508	508	508
Z	518	520	518	517	515	516	506	508	509	507	506	506

8. Um experimento foi executado por seis máquinas e dez operários, de modo que cada operário use cada máquina apenas uma vez. Complete o quadro a seguir e faça as conclusões ao nível de 5%.

Fonte de variação	*Soma dos quadrados*	*G.L.*	*Quadrados médios*	*Teste F*
Entre máquinas	904			
Entre operários	2.334			
Resíduo				
Total	5.832			

9. Aplique a regra de Scheffé para descobrir as médias diferentes, considerando os dados do Exercício 7.

RESPOSTAS DOS EXERCÍCIOS

SÉRIE I – CAPÍTULO 1

1. a) $S = \{1C, 2C, 3C, 4C, 5C, 6C, 1K, 2K, 3K, 4K, 5K, 6K\}$

   b) $A = \{K2, K4, K6\}$
   $B = \{C1, C3, C5\}$
   $C = \{3C, 6C, 3K, 6K\}$

   c) I) $\bar{B} = \{2C, 4C, 6C, 1K, 2K, 3K, 4K, 5K, 6K\}$
   II) $A \cup B = \{K2, K4, K6, C1, C3, C5\}$
   III) $B \cap C = \{3C\}$
   IV) $\overline{A \cup B} = \{K1, K3, K5, C2, C4, C6\}$

   d) A e B são mutuamente exclusivos

2. a) $\frac{1}{2}$ b) $\frac{3}{4}$ c) 0 d) $\frac{3}{4}$ e) 1

3. a) $\frac{7}{12}$ b) $\frac{3}{4}$ c) $\frac{5}{12}$

4. a) $\frac{1}{2}$ b) $\frac{1}{13}$ c) $\frac{7}{8}$ d) $\frac{2^n - 1}{2^n}$ e) $\frac{1}{17}$ f) $\frac{13}{102}$

5. a) $\frac{1}{5}$ b) $\frac{1}{10}$ c) $\frac{3}{10}$ d) 0,24

6. $\frac{4}{13}$

7. a) $\frac{1}{12}$ b) $\frac{1}{9}$ c) $\frac{5}{12}$

8. $\frac{4}{45}$

9. a) $\frac{7}{8}$ b) $\frac{5}{8}$ c) $\frac{3}{4}$

10. a) $\frac{3}{8}$ b) $\frac{7}{8}$ c) $\frac{91}{120}$ d) $\frac{1}{8}$

11. a) $\frac{4}{33}$ b) $\frac{5}{11}$ c) $\frac{31}{33}$

12. $\frac{5}{22}$

13. $$\frac{\binom{N_V}{n_V}\binom{N_A}{n_A}\binom{N_P}{n_P}}{\binom{N}{n}}$$

SÉRIE II – CAPÍTULO 1

1. a) $\frac{7}{12}$ b) $\frac{3}{4}$ c) $\frac{1}{2}$ d) 1

2. $\frac{5}{8}$; $\frac{5}{6}$

3. $$\frac{365.364 \ldots (365 - r + 1)}{365^r}$$

4. a) $\frac{28}{75}$ b) $\frac{1}{6}$ c) $\frac{1}{50}$

5. $2p^2 - p^4$

6. $\frac{35}{108}$

7. $\frac{5}{9}$

8. $\frac{9}{40}$

9. $\frac{5}{12}$

10. $\left(\frac{x}{x+y}\right)\left(\frac{z+1}{z+v+1}\right)+\left(\frac{y}{x+y}\right)\left(\frac{z}{z+v+1}\right)$

11. $\frac{1}{30}; \frac{1}{30}; \frac{3}{4}$

12. I) $\frac{3}{8}$ II) $\frac{19}{40}$ III) $\frac{9}{19}$

13. a) $\frac{3}{20}$ b) $\frac{3}{10}$ c) $\frac{9}{10}$ d) $\frac{9}{20}$

14. $\frac{4}{7}$

15. a) $\frac{7}{12}$ b) $\frac{2}{5}$

16. I) $\frac{41}{72}$ II) $\frac{13}{36}$

17. I) $\frac{20}{41}$ II) $\frac{3}{13}$

18. I) $\frac{1}{2}\left[\frac{x}{x+y}+\frac{z}{z+v}\right]$ II) $\left(\frac{xz+x+yz}{(x+y)(z+v+1)}\right)$

19. $\frac{x!y!}{(x+y)!}$

20. a) 1/6 b) 1/15 c) 1 d) 0 e) 0

f) 0 g) 2/15 h) 1/3 i) 0 j) 0

k) 0 l) 0

21. $\frac{21}{46}; \frac{25}{46}$

22. $\frac{10}{19}$

23. $\frac{4}{19}$

24. $\frac{25}{39}$

25. $\frac{8}{35}$

SÉRIE III – CAPÍTULO 1

1. a) $\frac{1}{8}$ b) $\frac{3}{8}$ c) $\frac{3}{8}$ d) $\frac{7}{8}$ e) $\frac{1}{8}$

2. a) $\frac{1}{6}$ b) $\frac{5}{6}$ c) $\frac{5}{12}$ d) $\frac{1}{2}$ e) $\frac{1}{5}$ f) $\frac{1}{6}$ g) 0

3. $\frac{29}{35}$

4. $\frac{2}{3}$

5. a) $\frac{2}{3}$ b) $\frac{7}{15}$ c) $\frac{1}{5}$ d) $\frac{5}{7}$ e) $\frac{3}{5}$

6. a) X e Y são independentes.

   b) M e N *não* são independentes.

7. a) 60% b) 26% c) 65% d) 21% e) 40% f) 65%

SÉRIE I – CAPÍTULO 2

1. a) Tabelas

I)

w	–5	–4	–3	–2	–1	0	1	2	3	4	5
$P(w)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

II)

A	2	4	6	8	10	12
$P(A)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

III)

z	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
$P(z)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

IV)

B	1	2	3	4	5	6
$P(B)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

b) Função Repartição de B

para $B < 1$ $F = 0$

para $1 \leqslant B < 2$ $F = \frac{1}{36}$

para $2 \leqslant B < 3$ $F = \frac{4}{36}$

para $3 \leqslant B < 4$ $F = \frac{9}{36}$

para $4 \leqslant B < 5$ $F = \frac{16}{36}$

para $5 \leqslant B < 6$ $F = \frac{25}{36}$

para $B \geqslant 6$ $F = 1$

c) I) $\frac{3}{4}$ II) $\frac{5}{9}$ III) $\frac{1}{2}$ IV) $\frac{5}{18}$ V) $\frac{1}{18}$ VI) $\frac{1}{4}$

VII) 0 VIII) $\frac{1}{6}$ IX) $\frac{7}{36}$ X) 0 XI) $\frac{1}{2}$ XII) $\frac{5}{6}$

2. a) $K = \frac{105}{176}$ b) $\frac{21}{176}$

3. a)

z	0	1	2	3	4	5	6	7	8	9	10	11	12
$P(z)$	$\frac{1}{28}$	$\frac{1}{28}$	$\frac{2}{28}$	$\frac{2}{28}$	$\frac{3}{28}$	$\frac{3}{28}$	$\frac{4}{28}$	$\frac{3}{28}$	$\frac{3}{28}$	$\frac{2}{28}$	$\frac{2}{28}$	$\frac{1}{28}$	$\frac{1}{28}$

c) $\frac{5}{14}$ d) $\frac{11}{14}$

4. a)

X	0	1	2	3
$P(x)$	$\frac{2}{42}$	$\frac{15}{42}$	$\frac{20}{42}$	$\frac{5}{42}$

b) $F(x) = 0$ para $x < 0$

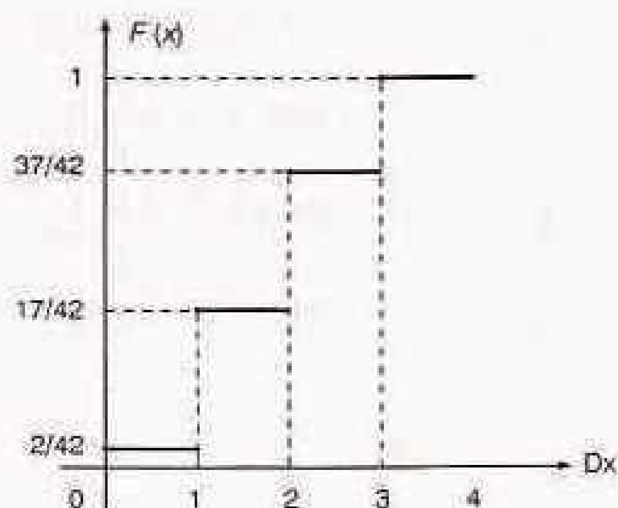
$F(x) = \frac{2}{42}$ para $0 \leq x < 1$

$F(x) = \frac{17}{42}$ para $1 \leq x < 2$

$F(x) = \frac{37}{42}$ para $2 \leq x < 3$

$F(x) = 1$ para $x \geq 3$

c)



d) I) $\frac{37}{42}$ II) $\frac{2}{42}$ III) $\frac{25}{42}$ IV) 0

V) $\frac{5}{42}$ VI) 1 VII) 1

e) $\frac{37}{42}$; 1; $\frac{4}{42}$; 1; $\frac{37}{42}$; $\frac{17}{42}$; 1; 0

SÉRIE II – CAPÍTULO 2

1. Para $x < 0$ $F(x) = 0$

   Para $0 \leqslant x < 1$ $F(x) = \frac{x(3 - x^2)}{2}$

   Para $x \geqslant 1$ $F(x) = 1$

2. Para $x < 0$ $F(x) = 0$

   Para $x \geqslant 2$ $F(x) = 1$

   Para $0 \leqslant x < 2$ $F(x) = \frac{x^2}{4}$

3. a) $\frac{1}{12}$ b) $\frac{1}{3}$

4. a) $\frac{1}{6}$

5. a) $K = 6$ b) $\begin{cases} F(x) = 0 & \text{para } x < 0 \\ F(x) = 3x^2 - 2x^3 & \text{para } 0 \leq x < 1 \\ F(x) = 1 & \text{para } x \geq 1 \end{cases}$ c) $\frac{1}{2}$

6. a) $A = \frac{1}{500^2}$ b) $\frac{3}{4}$

7. a) $\frac{1}{2}$ b) 0 c) 0

SÉRIE III – CAPÍTULO 2

1. a)

X	1	2
$P(X)$	0,6	0,4

Y	–2	–1	4	5
$P(Y)$	0,3	0,3	0,1	0,3

b) 1,4; 1; 0,9 c) –0,5 d) 0,49; 3,1 e) –0,3

f) Não pois, por exemplo, $P(x = 1, y = -2) \neq P(x = 1) \cdot P(y = -2)$

2. a)

M \ N	5	10	12
1	0,18	0,30	0,12
3	0,12	0,20	0,08

b) 1,8; 8,9 c) $\sqrt{0{,}96}$; $\sqrt{7{,}09}$

d) é igual a zero, pois M e N são independentes.

3. a) $g(x) = \frac{3}{2}x^2 + x$ $h(y) = \frac{3}{2}y^2 + y$

b) $\frac{17}{24}$; $\frac{17}{24}$ c) $\frac{139}{2880}$; $\frac{139}{2880}$ d) $\frac{39}{128}$ e) $\rho = -0{,}036$

4. a) $g(x) = x + \frac{1}{2}$ $h(y) = y + \frac{1}{2}$

b) $\frac{7}{12}$ c) $\frac{7}{12}$ d) $-\frac{1}{144}$

5. a) $\frac{1}{\ell n2}$ b) $\frac{1}{\ell n2}$ c) $\sqrt{2}$ d) 1 e) $\frac{3}{2\ell n2} - \frac{1}{\ell n^2 2}$

6. a) $K = 12$ b) $\frac{3}{5}$ c) $4md^3 - 3md^4 = 0,5$ d) $\frac{1}{25}$

7. a) 0,9 b) –2 c) 0,5 d) 6,29

8. a) $\frac{3}{4}$ b) $\frac{1}{\sqrt[3]{2}}$ c) $\frac{3}{80}$

10. $\mu = 700$ e $\sigma = 22,36$

11. a) $\frac{1}{18}$

b)

Y \ X	1	2	3
1	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{6}$
2	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$

c) $\frac{1}{3}$ d) $\frac{1}{2}$ e) $\frac{1}{6}$ f) $\frac{2}{3}$ g) $\frac{7}{3}$

h) $\frac{2}{9}$ i) $\frac{35}{9}$ j) 0 i) 0

12. a) $\frac{3}{2}$ b) $\frac{1}{4}$ c) $\frac{29}{64}$ d) $\frac{5}{16}$ e) não f) $-\frac{1}{64}$ g) –0,21

13. sim; 409,27

14. $\mu = 2$; $m_0 = 2$, $\sigma = 0,71$

15. $\mu = 340$; $\sigma = 2,67$

16. $\mu_{(L)} = 28,5$; $\sigma_{(L)} = 6,21$

SÉRIE I – CAPÍTULO 3

Distribuição Binomial

1. a) $\frac{105}{512}$ b) $\frac{1.013}{1.024}$ c) $\frac{1}{1.024}$ d) $\frac{1.023}{1.024}$ e) $\frac{193}{256}$

2. $\frac{15}{64}$

3. a) 20 b) 80 c) 20

4. $1 - \left(\frac{5}{6}\right)^n$

5. a) $\frac{80}{243}$ b) $\frac{242}{243}$ c) $\frac{64}{81}$

6. a) $\frac{80}{243}$ b) $\frac{64}{729}$

7. $\binom{100}{70}\left(\frac{1}{2}\right)^{100}$

8. a) 5 b) $\frac{2}{3}$ e $\frac{1}{3}$ c) $\frac{10}{3}$ d) $\frac{10}{9}$ e) $\frac{242}{243}$ f) $\frac{200}{243}$

9. a) $(0,95)^{100}$ b) $\binom{100}{3}(0,05)^3(0,95)^{97}$

c) $1 - (0,05)^{100} - 100\,(0,95)\,(0,05)^{99}$

Distribuição Multinomial

11. $\frac{35}{5832}$

12. 0,0324

13. a) 0,0216 b) 0,1296

Distribuição de Poisson

14. a) 0,8784 b) 0,2020
15. a) 0,168 b) 0,5767
16. a) 0,2241 b) 0,658
17. 0,1952
18. a) 0,0183 b) 0,0732 c) 0,1464 d) 0,9085
19. a) 0,449 b) 0,1437
20. a) 0,270 b) 0,180

SÉRIE I – CAPÍTULO 4

Distribuição Uniforme

1. a) $\frac{1}{3}$ b) $\frac{1}{2}$ c) 0 d) 2,5 e) $\frac{3}{4}$
3. a) 3 c) $\frac{5}{4}$

Distribuição Normal

4. a) 0,4251 b) 0,3023 c) 0,9104 d) 0,2064 e) 0,1401
   f) 0,7454 g) 0,3830
5. a) 1 b) 0,8665 c) 0,0132 d) 0 e) 776 dias
6. a) 380 b) 389
7. 88,5 e 55
8. a) 0,8413 b) 0,7745
9. a) 0,0013 b) 0,7210
10. a) 0,3530 b) $128,75 \leq X \leq 231,25$
11. 0,0197
12. a) 0,0643 b) 0,6480
13. a) 0,2981 b) 0,0409
14. $\mu = 29,03$; $\sigma^2 = 73,44$
15. $Y = N\,(13;\,38)$
16. a) 7,3% b) 0,2266
17. E_1
18. a) 425,60 b) 0,1112
19. a) 68,26% b) 95,44% c) 99,74% d) 70,62%
    e) 100% aproximadamente
21. a) 4,88 b) 30,4 c) 21,04 d) 124,5
    e) 2,99 f) 85,37 g) 30
22. a) −1,64 b) −1,17 c) −0,39 d) 0
    e) 0,25 f) 0,67 g) 1,28 h) −0,58
    i) −0,39 j) 0,31 l) 1,41

Distribuição Exponencial

23. a) 0,6321 b) 0,3679 c) 1.000

24. a) 0,2212 b) 0,0323 c) 0 d) 0,4724

SÉRIE II – CAPÍTULO 4

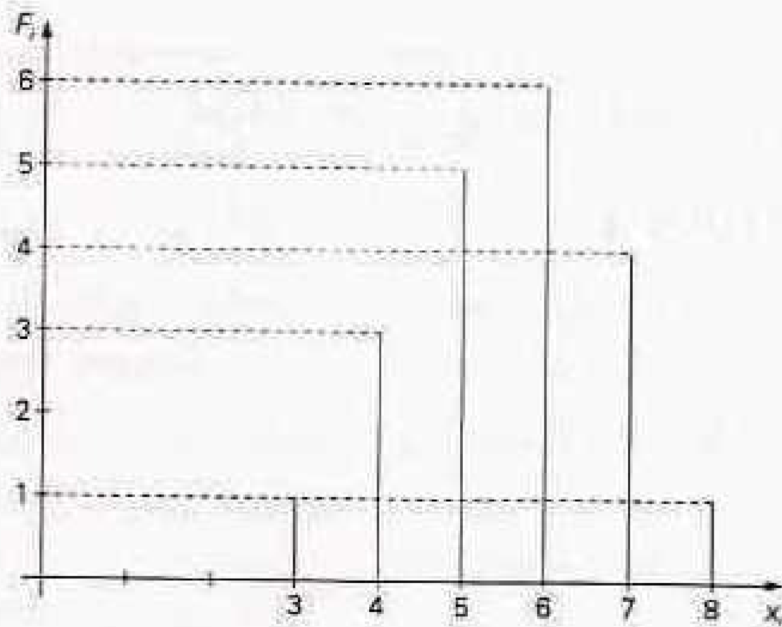
1. $\mu(\chi^2_{23}) = 23$ $\sigma^2(\chi^2_{23}) = 46$ $\sigma(\chi^2_{23}) = 6,78$
   $Md(\chi^2_{23}) = 22,3$ $Q_3 = 27,1$
2. $\chi^2_{sup} = 13,4$ $\chi^2_{inf} = 3,49$
3. $\mu(t_{23}) = 0$ $\sigma^2(t_{23}) = 1,095$ $\sigma(t_{23}) = 1,046$
   $Q_1 = -0,68531$ $P_5 = -1,7139$ $Mo = 0$
4. $-1,1848$ 2,0860
5. $\mu = 1,25$ $\sigma^2 = 1,042$ $\sigma = 1,021$ $Mo = 0,625$
   0,2985 4,07
6. a) 95,31 b) −0,39 c) 0,84 d) 0,9340
   e) R$ 2.019,2 f) 29,3 g) 8,55 h) 15,5
   i) 34,4 j) 0,925 l) 0,72669 m) 0,025
   n) 0,005 o) 0,8725 p) 1,7033 q) 0,745
   r) 0,2857 s) 3,50 t) 0,90 u) 0,05

SÉRIE II – CAPÍTULO 5

1. a)
   b)
   c)

X_i	F_i	f_i	F_{ac}
3	1	0,05	1
4	3	0,15	4
5	5	0,25	9
6	6	0,30	15
7	4	0,20	19
8	1	0,05	20

d)

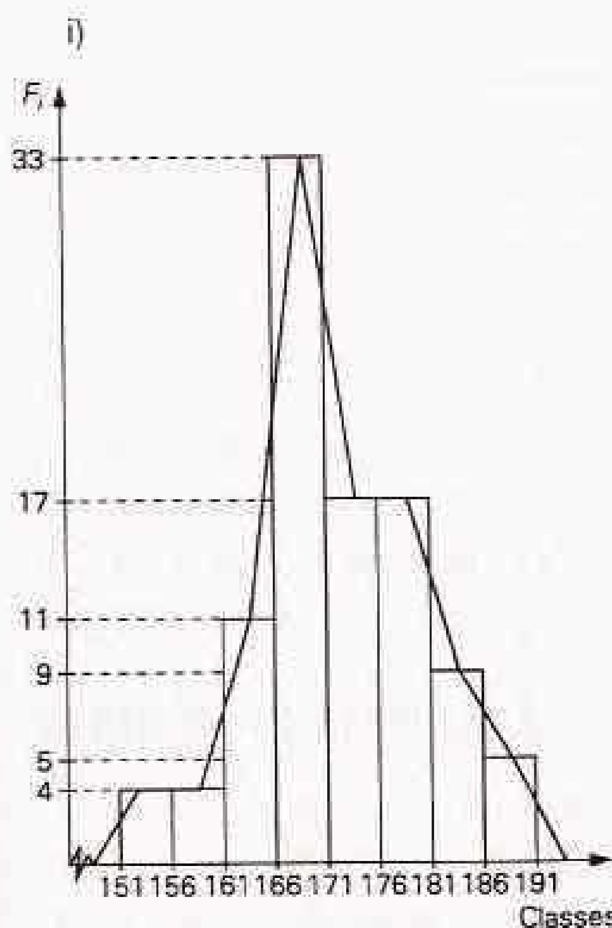


e) 5 f) 55%

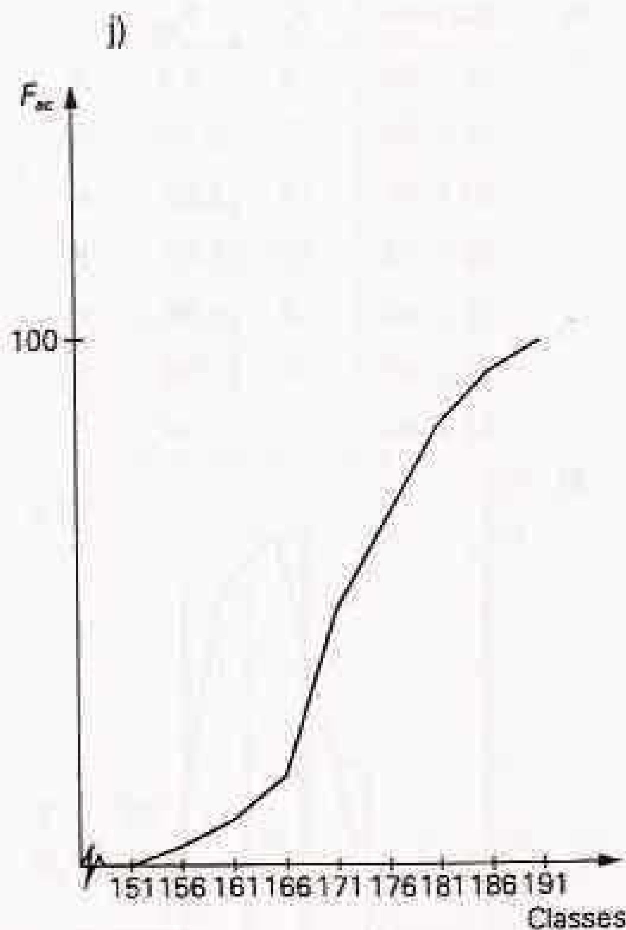
2. a) $R = 39$ b) $K = 8$ c) $h = 5$ d) e) f) g) h)

Limites das classes	F_i	X_i	f_i	F_{ac}
151 ⊢ 156	4	153,5	0,04	4
156 ⊢ 161	4	158,5	0,04	8
161 ⊢ 166	11	163,5	0,11	19
166 ⊢ 171	33	168,5	0,33	52
171 ⊢ 176	17	173,5	0,17	69
176 ⊢ 181	17	178,5	0,17	86
181 ⊢ 186	9	183,5	0,09	95
186 ⊢ 191	5	188,5	0,05	100

i)



j)



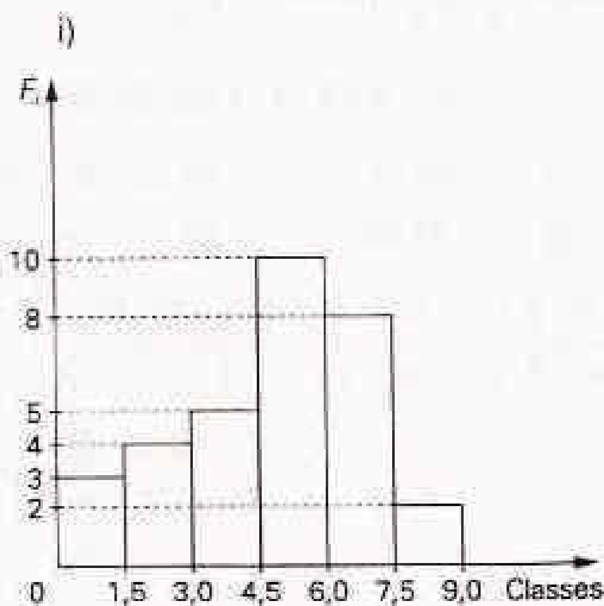
3. a) 0 – 0 – 1 – 1,5 – 2 – 2 – 2,5 – 3,5 – 3,5 – 4 – 4 – 4 – 4,5 – 4,5 – 4,5 – 5 – 5 – 5 – 5 – 5 – 5,5 – 5,5 – 6 – 6 – 6 – 6,5 – 6,5 – 7 – 7 – 7 – 8 – 8,5

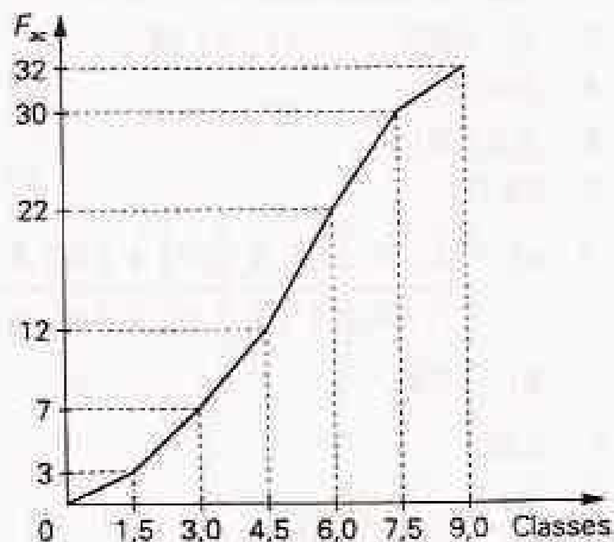
b)

Classes	**0 ⊢ 1,5**	**1,5 ⊢ 3,0**	**3,0 ⊢ 4,5**	**4,5 ⊢ 6,0**	**6,0 ⊢ 7,5**	**7,5 ⊢ 9,0**
F_i	3	4	5	10	8	2

c) 8,5 e 0 d) 8,5 e) 28,185% f) 3 g) 4,5 h) 3,75

i)

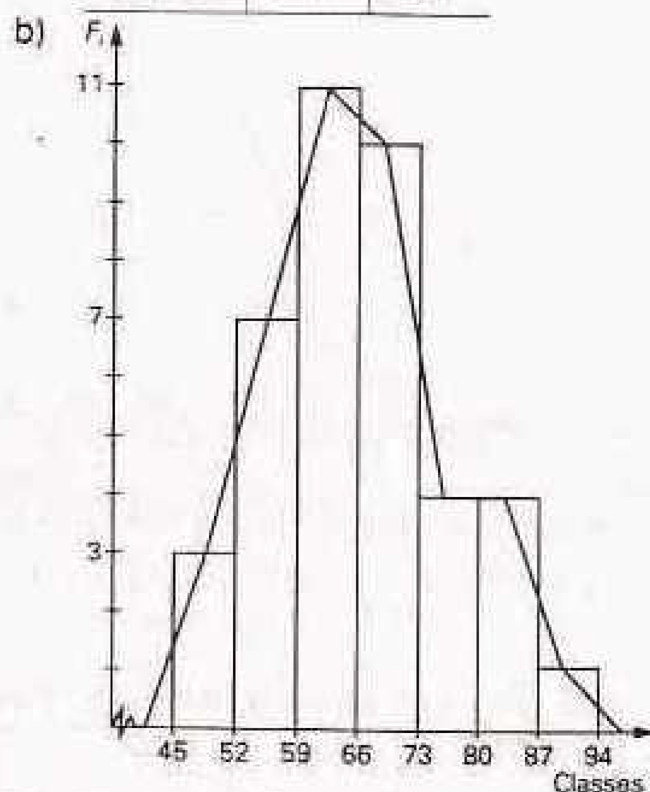


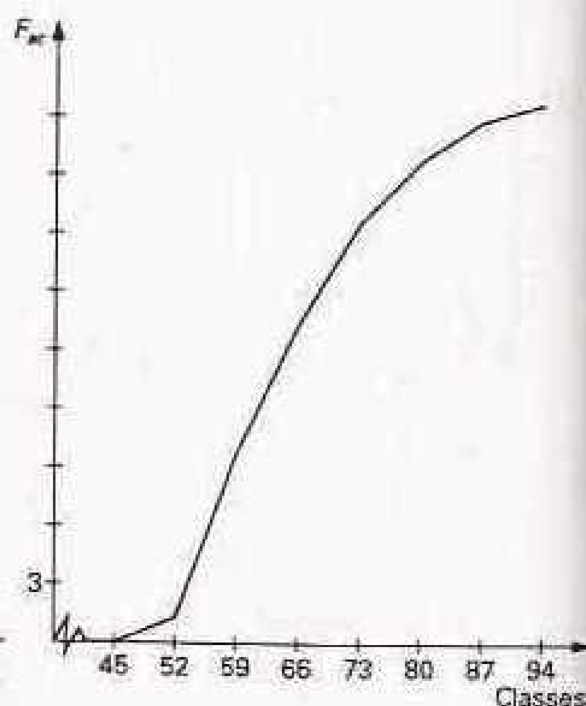


4. a)

Classes	F_i	F_{ac}
45 ⊢ 52	3	3
52 ⊢ 59	7	10
59 ⊢ 66	11	21
66 ⊢ 73	10	31
73 ⊢ 80	4	35
80 ⊢ 87	4	39
87 ⊢ 94	1	40

b)





SÉRIE III – CAPÍTULO 5

1. a) 4 b) 9 c) 3,305 d) 79,43
2. Não foi aprovado: $\bar{x} = 4{,}875$
3. a) 6,82 b) 11,59 c) 4 d) 9,03 e) 87,88
4. 164,93
5. $ 5.820
6. 78,16
7. a)

X_i	0	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25
F_i	2	4	4	4	4	4	4	4	5	4	2	2	1	2	2	2

   b) 7,24
8. 5,85
9. a) b)

Classes	F_i	X_i
15 ├ 20	10	17,5
20 ├ 25	4	22,5
25 ├ 30	12	27,5
30 ├ 35	24	32,5
Σ	50	

c) 27,5

10. a) 10,95 b) 5,22 c) 9,29
11. a) 8,12 b) 3,53
12. 355,93

SÉRIE IV – CAPÍTULO 5

1. I) 4 II) 5 III) 8 IV) 87
2. I) 4 II) 77 III) 13 IV) 235
3. I) 6,63 II) 28,35
4. I) 7 II) 43
5. I) 80 II) 3,5
6. I) 14,5 II) 26,25
7. I) 8,8; 9,03; 6,86 II) 33,6; 42,32; 50
8. a) 1,17 b) 1 c) 0 d) 34%
9. a)

X_i	1	2	3	4	5	6	7	8	10
F_i	1	1	5	6	3	2	3	2	1

b) 4,83; 4; 4

10. a) 22,99 b) 21,85 c) 20,18 d) 18,59
e) 17,68 f) 29,48 g) 74%
11. a)

X_i	0	1	2	3	4	5	6	7	8	9
F_i	2	3	4	8	1	5	5	7	3	2

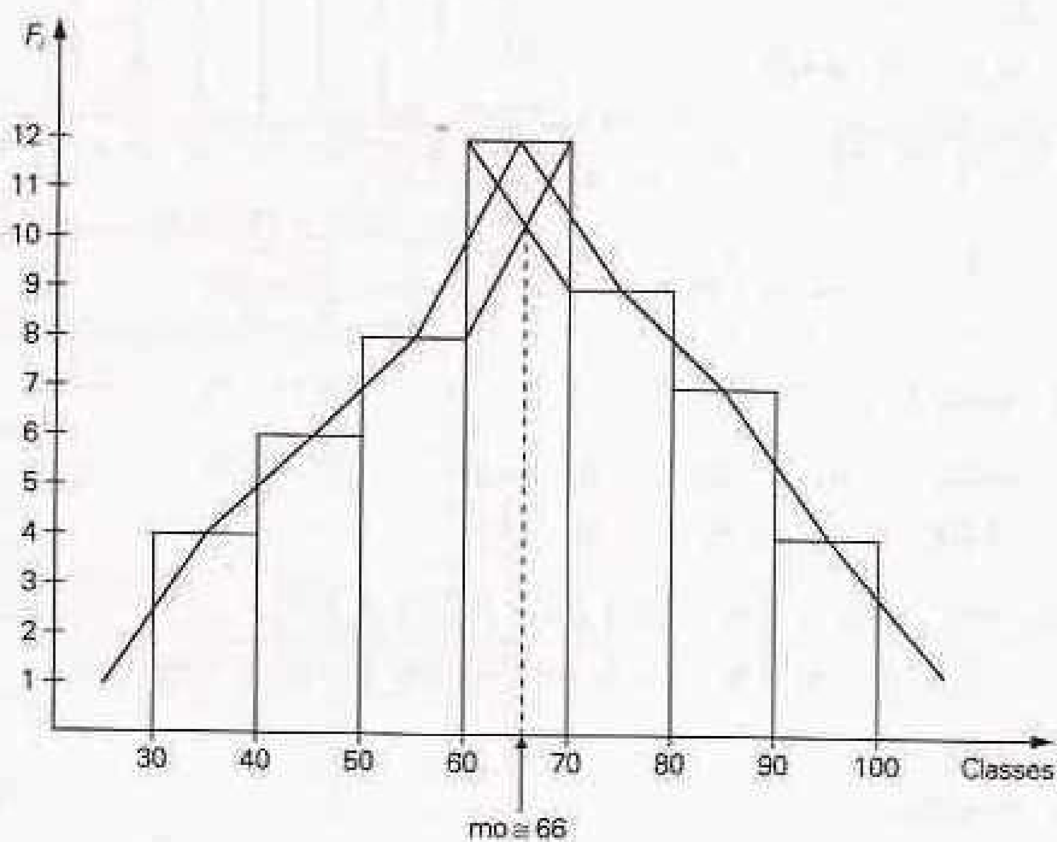
b) 4,63
c) 7; moda
d) 5

12. a) 65 b) 7 c) 10

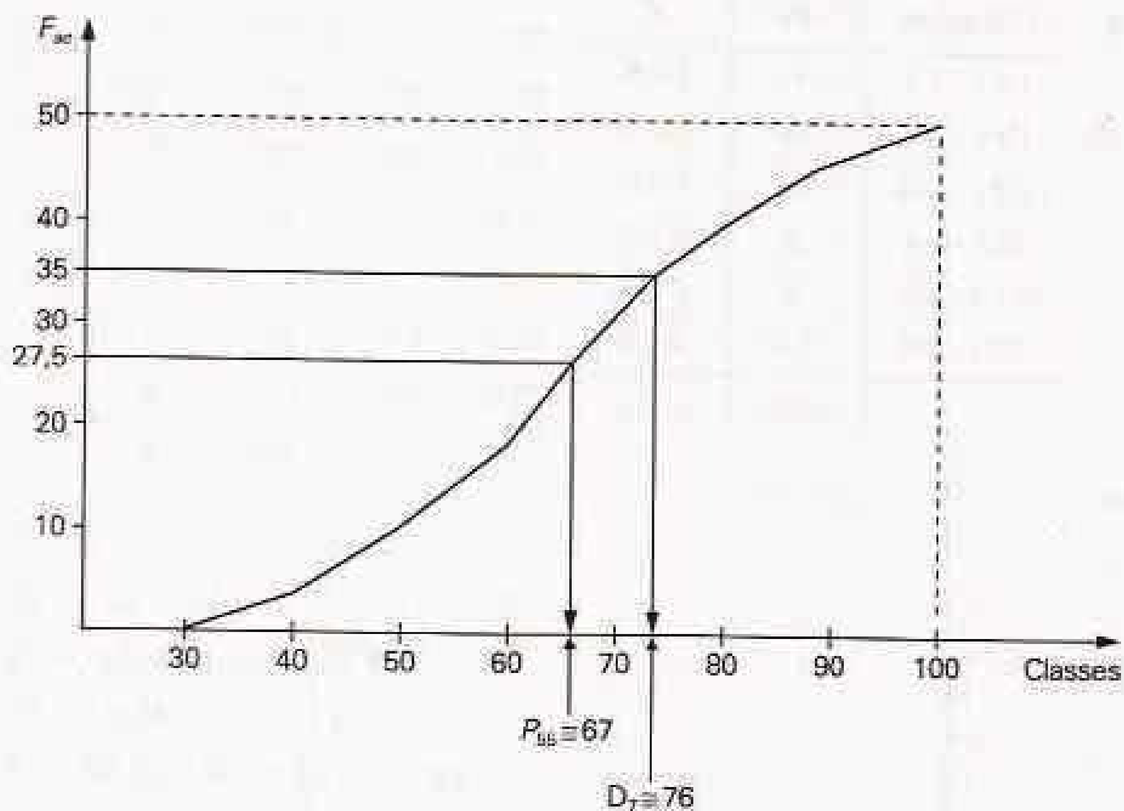
d) e) f) g) h)

Classes	F_i	f_i	X_i	F_{ac}
30 ⊢ 40	4	$\frac{2}{25}$	35	4
40 ⊢ 50	6	$\frac{3}{25}$	45	10
50 ⊢ 60	8	$\frac{4}{25}$	55	18
60 ⊢ 70	12	$\frac{6}{25}$	65	30
70 ⊢ 80	9	$\frac{9}{50}$	75	39
80 ⊢ 90	7	$\frac{7}{50}$	85	46
90 ⊢ 100	4	$\frac{2}{25}$	95	50
Σ	50	1		

i) j) m)



k) n) o) p) l) 65,8



SÉRIE V – CAPÍTULO 5

1. a) 10 b) 3,02 c) 13,81

2. a)

X_i	5	6	7	8	9
F_i	3	4	6	3	2

b) 4 c) 0,98 d) 1,47 e) 1,21 f) 18%

3. 5,84

4. a) 11 b) 175 c) 13,23
d) 20% e) 0,38 f) 0,260

5. a) 53,5 b) 45 c) 13%
d) $A_s = 0{,}21$, a distribuição não é simétrica.
e) $K = 0{,}260$, a distribuição não é mesocúrtica.

6. $\bar{x} = 55{,}5$; $s^2 = 126$; $s = 11{,}22$; $cv = 20\%$; $A_s = -0{,}045$;
$K = 0{,}275$

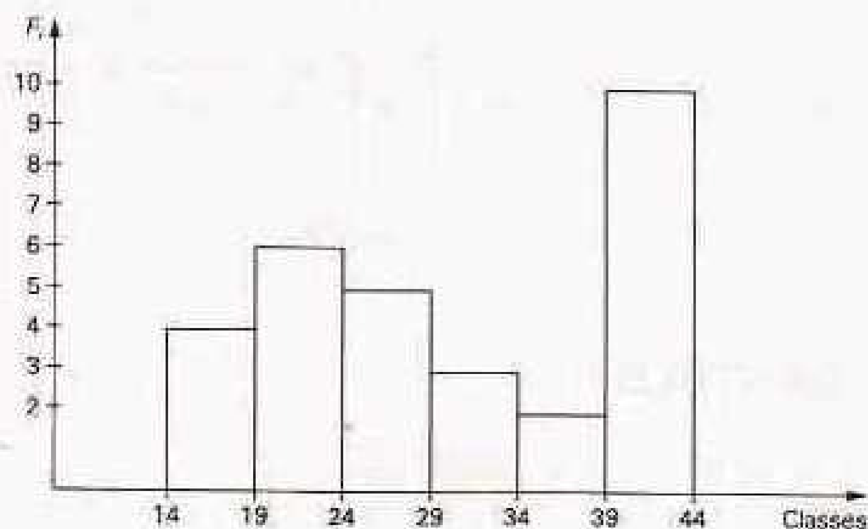
7. 50%

8. a) A b) A

9. a)

Classes	F_i	X_i
14 ⊢ 19	4	16,5
19 ⊢ 24	6	21,5
24 ⊢ 29	5	26,5
29 ⊢ 34	3	31,5
34 ⊢ 39	2	36,5
39 ⊢ 44	10	41,5
	30	

b)



c) 30,33 e 9,53

10. a) 295.600 b) $A_s = 0{,}32$ Sim c) $K = 0{,}227$ Sim

SÉRIE VI – CAPÍTULO 5

1. $\bar{x} = 1{,}58$; $\sigma = 0{,}286$
2. a) 171,59 b) 171,82 c) 172,67 d) 3,99
   e) −0,22
3. $ 37,08

4.

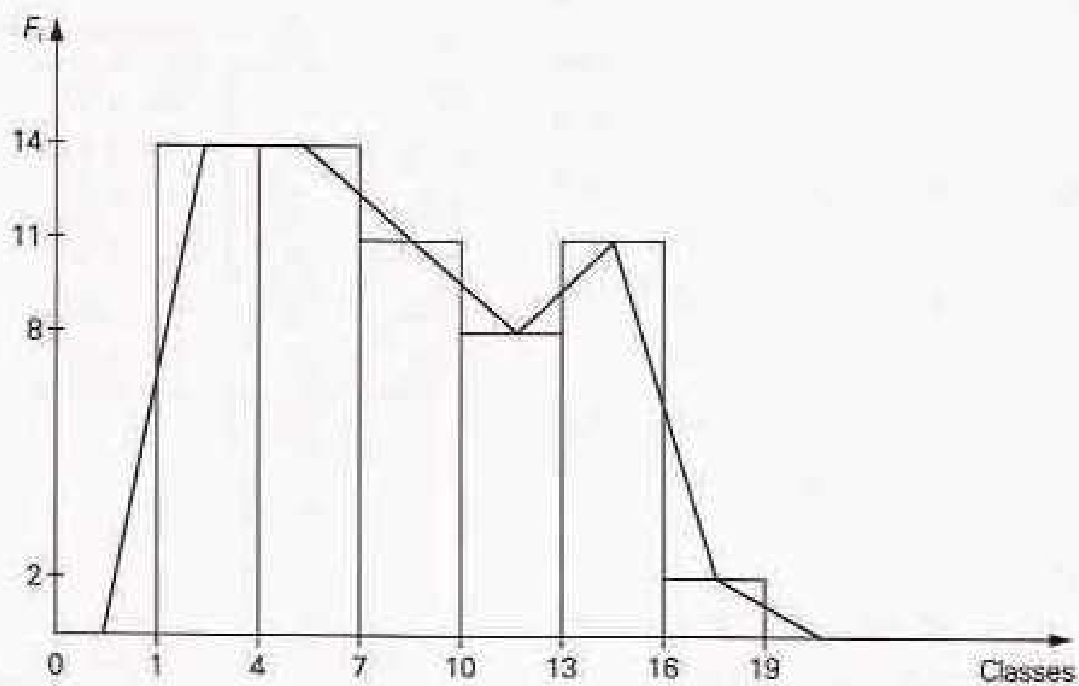
X_i	F_i	F_{ac}	f_i
1	4	4	0,04
2	8	12	0,08
3	18	30	0,18
4	27	57	0,27
5	15	72	0,15
6	11	83	0,11
7	10	93	0,10
8	7	100	0,07
Σ	100		1

5. $F_i = 7$
6. $Q_1 = 1{,}21$; $D_7 = 3{,}1$; $P_{73} = 3{,}25$
7. $s^2 = 422{,}68$; $mo = 55{,}56$
8. $\sigma^2 = 3{,}05$; $\sigma = 1{,}75$
9. $\bar{x} = 30.003{,}69$; $s^2(x) = 7{,}43$
10. $A_s = -0{,}11$ ∴ a distribuição é assimétrica negativa
    $K = 0{,}2718$ ∴ a distribuição é platicúrtica
11. a) 44% b) 45 c) 15,13 d) 45
    e) A equipe 2, pois $\sigma = 15$ e c.v. = 34%

12. a)

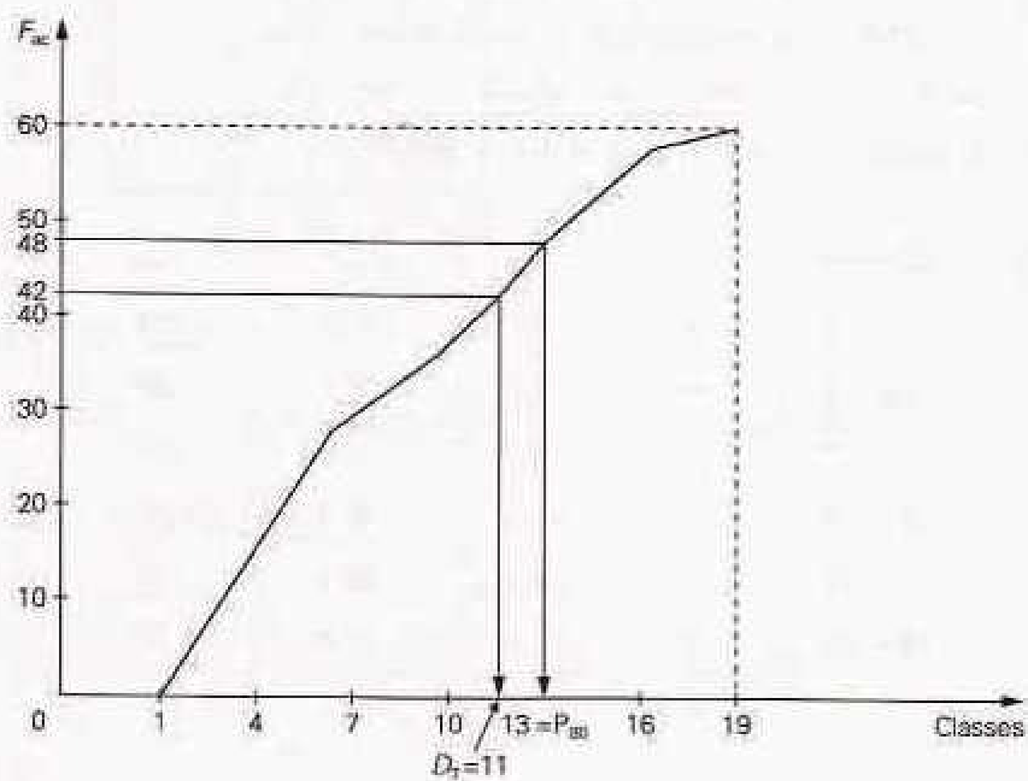
Classes	F_i	X_i	$\lvert d_i \rvert F_i$	F_{ac}
1 ⊢ 4	14	2,5	79,8	14
4 ⊢ 7	14	5,5	37,8	28
7 ⊢ 10	11	8,5	3,3	39
10 ⊢ 13	8	11,5	26,4	47
13 ⊢ 16	11	14,5	69,3	58
16 ⊢ 19	2	17,5	18,6	60
Σ	60		235,2	

b)



F_r
14
11
8
2
0
1
4
7
10
13
16
19
Classes

c)



F_{ac}
60
50
48
42
40
30
20
10
0
1
4
7
10
13 = P_{80}
16
19
Classes
D_7 = 11

d) 8,2 e) 7,55 f) 12,25 g) 6,14 h) 7,05
i) 4,21 j) 3,92 l) 21,26 m) 4,61 n) 56%
o) A_s = 0,49, não
p) K = 0,3185, não
q) vide gráfico da F_{ac}

SÉRIE VII – CAPÍTULO 5

1. b
2. b
3. c
4. d
5. b
6. a
7. d
8. b
9. d
10. a
11. a
12. d
13. d
14. a
15. d
16. d
17. b
18. a
19. b
20. b
21. d
22. c
23. d
24. b
25. c
26. c
27. d
28. c
29. a
30. b

SÉRIE I – CAPÍTULO 6

1. a) $\mu(x) = 3,5$ b) $\sigma(x) = 1,1180$
   c) $\mu(\bar{x}) = 3,5$ d) $\sigma(\bar{x}) = 0,7906$

5. 52.000
6. 19.888

SÉRIE I – CAPÍTULO 7

1. a) $n = 33$
4. $n = 400$
5. $n = 399$. Comparando-se os resultados de 4 e 5 verifica-se que uma população de 200.000 dá aproximadamente o resultado de uma população infinita.

SÉRIE I – CAPÍTULO 8

1. O intervalo [4,81 ; 5,59] contém a média populacional com 90% de confiança;
   O intervalo [4,73 ; 5,67] contém a média populacional com 95% de confiança;
   O intervalo [4,58 ; 5,82] contém a média populacional com 99% de confiança.
2. O intervalo [25,76 ; 28,00] contém a média populacional com 95% de confiança;
   O intervalo [25,94 ; 27,82] contém a média populacional com 90% de confiança.
3. O intervalo [172,06 cm ; 177,94 cm] contém a verdadeira altura média dos alunos com 95% de confiança.
4. O intervalo [104,2 ; 115,8] contém a média populacional com 90% de confiança;
   O intervalo [102,84 ; 117,16] contém a média populacional com 95% de confiança.
5. O intervalo [9,164 ; 12,586] contém a média populacional com 97,5% de confiança;
   O intervalo [9,993 ; 11,757] contém a média populacional com 75% de confiança.
6. Satisfaz, pois o intervalo [288,33 kg ; 304,93 kg] contém a média da população com 95% de confiança.
7. a) $\bar{x} = 13,13$ $\qquad s^2 = 2,05$
   b) O intervalo [12,60 ; 13,66] contém a média populacional com 95% de confiança.
8. Os limites de confiança para a média são: [26,38 seg ; 32,02 seg] com 90%.
9. a) O intervalo [7,26 ; 13,58] contém a média populacional com 95% de confiança;
   b) O intervalo [22,00 ; 24,55] contém a média populacional com 95% de confiança;
   c) O intervalo [7,98 ; 12,67] contém a média populacional com 95% de confiança.
10. a) O intervalo [0,32 ; 3,13] contém a variância populacional com 90% de confiança;
    b) O intervalo [2,25 ; 8,13] contém a variância populacional com 90% de confiança.
11. Os limites de confiança a 80% para a variância são [1,38 ; 4,86].

12. O intervalo [0,85 ; 3,95] contém a variância populacional com 95% de con

8. Como $\chi^2_{cal} = 5{,}57$, não se pode rejeitar a hipótese de que a variância populacional é 4, ao nível de 1%.

9. Como $Z_{cal} = 0{,}89$, não se pode rejeitar a hipótese de que a proporção de eleitores democratas é 50%, ao nível de 5%.

10. a) Como $Z_{cal} = 0{,}33$, não se pode rejeitar, ao nível de 4%, a hipótese de que a proporção de fumantes seja de 80%.

    b) Como $Z_{cal} = 2{,}75$, rejeita-se H_0, concluindo-se, com risco de 2%, que a proporção dos que fumam cigarros com filtro é diferente de 70%.

    c) Como $Z_{cal} = 4{,}05$, rejeita-se H_0, concluindo-se, com risco de 1%, que a proporção de fumantes femininas é diferente de 40%.

11. Como $Z_{cal} = 2$, rejeita-se H_0 ($p = 0{,}50$), concluindo-se ao nível de 5% que a moeda não é honesta.

12. Como $Z_{cal} = 4{,}47$, rejeita-se H_0, concluindo-se, com risco de 4%, que a proporção é diferente de 0,5.

13. Como $Z_{cal} = 0{,}85$, não se pode rejeitar a hipótese de que a proporção de reprovados seja 40%, ao nível de 5%.

14. Como $F_{cal} = 1{,}46$, não se pode rejeitar, ao nível de 10%, a hipótese de que as variâncias são iguais.

16. Como $F_{cal} = 0{,}347$, não se pode rejeitar, ao nível de 10%, a hipótese de que as variâncias são iguais.

17. Como $Z_{cal} = 1{,}29$, não se pode rejeitar a hipótese de igualdade das médias, ao nível de 4%.

18. Como $t_{cal} = 1{,}1$, não se pode rejeitar a hipótese de igualdade das médias, ao nível de 5%.

20. Como $t_{cal} = 0{,}35$, não se pode rejeitar a hipótese de que a diferença valha $ 5.000,00, ao nível de 10%.

21. Como $Z_{cal} = -1{,}67$, não se pode rejeitar a hipótese de igualdade entre as proporções ao nível de 5%.

22. Como $Z_{cal} = -10$, rejeita-se H_0, concluindo-se, com risco de 5%, que os dois processos são diferentes.

23. Como $Z_{cal} = -2{,}86$, rejeita-se H_0, concluindo-se, com risco de 10%, que as proporções são diferentes.

SÉRIE I – CAPÍTULO 10

1. Como $\chi^2_{cal} = 2$, não se pode rejeitar a hipótese de honestidade da moeda, ao nível de 10%.
2. Como $\chi^2_{cal} = 1{,}6$, não se pode rejeitar a hipótese de honestidade do dado, aos níveis de 5% e 2,5%.
3. Como $\chi^2_{cal} = 0{,}56$, não se pode rejeitar a hipótese de honestidade dos dados, ao nível de 1%.
4. Como $\chi^2_{cal} = 19{,}28$, rejeita-se H_0, concluindo-se, ao nível de 5%, que os números não são aleatórios.
5. Como $\chi^2_{cal} = 16{,}92$, rejeita-se a hipótese de que as distribuições sejam iguais, ao nível de 2,5%.
6. Como $\chi^2_{cal} = 7{,}296$, não se pode rejeitar a hipótese de que o número de livros emprestados independe do dia da semana, ao nível de 1%.

SÉRIE II - CAPÍTULO 10

1. Como $\chi^2_{cal} = 6{,}11$, não se pode rejeitar a hipótese de que as variáveis sejam independentes, ao nível de 5%.
2. Como $\chi^2_{cal} = 1{,}52$, não se pode rejeitar a hipótese de haver homogeneidade aos níveis de 5% e 1%.
3. $c = 0{,}04 = 4\%$.
4. Como $\chi^2_{cal} = 5{,}81$, não se pode rejeitar a hipótese de independência entre as variáveis, ao nível de 1%.
5. Como $\chi^2_{cal} = 16{,}83$, rejeita-se H_0, concluindo-se, com risco de 5%, que há dependência entre as opiniões sobre a pena de morte e os partidos.

SÉRIE III - CAPÍTULO 10

1. a) (Sinais) Como $Z_{cal} = -0{,}75$, não se pode rejeitar a hipótese de que não houve diferença dos pesos, ao nível de 2,5%.

   a) (Wilcoxon) Como $Z_{cal} = -3{,}12$, rejeita-se H_0, concluindo-se, ao nível de 2,5%, que houve diferença.

   b) (Sinais) Como $Z_{cal} = -1{,}04$, não se pode rejeitar a hipótese de que não há diferença de consumo, ao nível de 2,5%.

   b) (Wilcoxon) Como $Z_{cal} = -1{,}0$, não se pode rejeitar a hipótese de que não há diferença de consumo, ao nível de 2,5%.

2. Como $Z_{cal} = 0{,}43$, não se pode rejeitar a hipótese de igualdade das médias ao nível de 1%.

3. Como $\chi^2_{cal} = 0{,}23$, não se pode rejeitar a hipótese de igualdade das medianas ao nível de 1%. As conclusões são as mesmas.

4. Como $Z_{cal} = -0{,}62$, não se pode rejeitar a hipótese de igualdade das marcas A e B, ao nível de 5%.

5. Como $H = 4{,}30$, não se pode rejeitar a hipótese de igualdade das vidas médias das 4 marcas, ao nível de 5%.

6. Como $H = 3{,}37$, não se pode rejeitar a hipótese de igualdade entre as vendas dos três shoppings ao nível de 2,5%.

SÉRIE I – CAPÍTULO 11

1. a) $\bar{x}_1 = 26{,}50$ $\quad \bar{x}_2 = 20$ $\quad \bar{x}_3 = 32{,}25$ $\quad \bar{x}_4 = 20{,}17$
   b) $\bar{x} = 24{,}45$
   c) 542,95
   d) 457,37
   e) 85,08
   f) Há diferença. $F_{cal} = 28{,}69$
   g) $\mu_1 \neq \mu_2$; $\mu_1 \neq \mu_3$; $\mu \neq \mu_4$;
   $\mu_2 \neq \mu_3$; $\mu_3 \neq \mu_4$

2. $F_{cal} = 1{,}92$. Não há diferença significativa entre as válvulas.
3. $F_{cal} = 1$. Não há evidência.
4. $F^C_{cal} = 6{,}43$
   $F^L_{cal} = 0{,}29$; só há entre as máquinas.
5. $F^C_{cal} = 5{,}83$. Há diferença devido à variedade do café.
   $F^L_{cal} = 0{,}65$. Não há diferença devido ao solo.
6. a) $F^L_{cal} = 9{,}03$. A distribuição interfere nas quantidades.
   b) $F^C_{cal} = 1.968{,}54$. O preço interfere nas quantidades.
   c) $F^I_{cal} = 0{,}92$. Não há efeito devido à interação.
7. a) Não são constantes. $F^C_{cal} = 41{,}68$
   b) São iguais. $F^L_{cal} = 2{,}02$
   c) Há interação: $F^I_{cal} = 3{,}78$

8. F_{cal} (op.) = 4,5

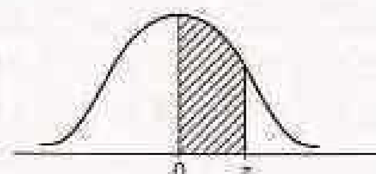
F_{cal} (máq.) = 3,14

Há diferença entre operários e entre máquinas.

9. $\mu_1 \neq \mu_2$; $\mu_1 \neq \mu_3$; $\mu_1 \neq \mu_4$; $\mu_2 \neq \mu_3$; $\mu_2 \neq \mu_4$

Tabela 1. Áreas de uma distribuição normal padrão

Cada casa na tabela dá a proporção sob a curva inteira entre $z = 0$ e um valor positivo de z. As áreas para os valores de z negativos são obtidas por simetria.



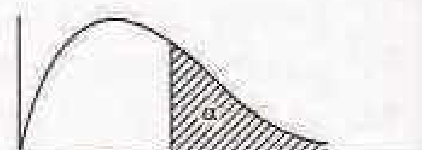
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767

(*continua*)

(continuação)

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Tabela 2. Distribuição de χ^2



φ \ α	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0001	0,0506	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6
3	0.0717	0,115	0,216	0,352	0,584	1,021	2,37	4,11	6,25	7,81	9,25	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3

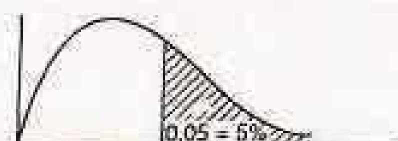
(continua)

(continuação)

φ \ α	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005
15	4,60	5,23	6,23	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8
16	5,14	5,80	6,91	7,96	8,31	11,9	15,3	19,4	23,5	26,3	28,4	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	53,4	35,7
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,5	42,8
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,2	33,1	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,5
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7

Para $\varphi > 30$ usar a aproximação: $\chi^2_x = \frac{1}{2}\left[\pm Z_\alpha + \sqrt{2\varphi - 1}\right]^2$

Tabela 3. Distribuição de F de Snedecor $\alpha = 5\%$



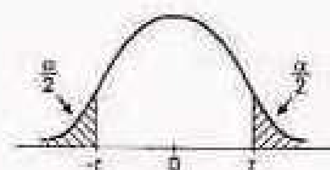
φ_2 \ φ_1	1	2	3	4	5	6	7	8	9	10	20	30	120	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	248,0	250,1	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,45	19,46	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,66	8,62	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,80	5,75	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,56	4,50	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,87	3,81	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,44	3,38	3,27	2,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,15	3,08	2,97	2,92
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	2,94	2,86	2,75	2,71

(continua)

(continuação)

φ_2 \ φ_1	1	2	3	4	5	6	7	8	9	10	20	30	120	∞
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,77	2,70	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,65	2,57	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,90	2,85	2,80	2,75	2,54	2,47	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,46	2,38	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,39	3,31	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,33	2,25	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,28	2,19	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,23	2,15	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,19	2,11	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,16	2,07	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,12	2,04	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,10	2,01	1,87	1,81
22	4,30	3,44	4,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,07	1,98	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,05	1,96	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,03	1,94	1,79	1,73
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	1,93	1,84	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,21	2,18	2,12	2,08	1,84	1,74	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,75	1,65	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,66	1,55	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,57	1,46	1,22	1,00

Tabela 4. Distribuição *t* de Student



φ \ α	0,50	0,25	0,10	0,05	0, 025	0,01	0,005
1	1,00000	2,4142	6,3138	12.706	25.542	63.657	127,32
2	0,81650	1,6036	2,9200	4,3127	6,2053	9,9248	14,089
3	0,76489	1,4226	2,3534	3,1825	4,1765	5,8409	7,4533
4	0,74070	1,3444	2,1318	2,7764	3,4954	4,6041	5,5976

(continua)

(*continuação*)

φ \ α	0,50	0,25	0,10	0,05	0, 025	0,01	0,005
5	0,72669	1,3009	2,0150	2,5706	3,1634	4,0321	4,7733
6	0,71756	1,2733	1,9432	2,4469	2,9687	3,7074	4,3168
7	0,71114	1,2543	1,8946	2,3646	2,8412	3,4995	4,0293
8	0,70639	1,2403	1,8595	2,3060	2,7515	3,3554	3,8325
9	0,70272	1,2297	1,8331	2,2622	2,6850	3,2498	3,6897
10	0,69981	1,2213	1,8125	2,2281	2,6338	3,1693	3,5814
11	0,69745	1,2145	1,7959	2,2010	2,5931	3,1058	3,4966
12	0,69548	1,2089	1,7823	2,1788	2,5600	3,9545	3,4284
13	0,69384	1,2041	1,7709	2,1604	2,5326	3,0123	3,3725
14	0,692	1,2001	1,7613	2,1448	2,5096	2,9768	3,3257
15	0,69120	1,1967	1,7530	2,1315	2,4899	2,9467	3,2860
16	0,69013	1,1937	1,7459	2,1199	2,4729	2,9208	3,2520
17	0,68919	1,1910	1,7396	2,1098	2,4581	2,8982	3,2225
18	0,68837	1,1887	1,7341	2,1009	2,4450	2,8784	3,1966
19	0,68763	1,1866	1,7291	2,0930	2,4334	2,8609	3,1737
20	0,68696	1,1848	1,7247	2,0860	2,4231	2,8453	3,1534
21	0,68635	1,1831	1,7207	2,0796	2,4138	2,8314	3,1352
22	0,68580	1,1816	1,7171	2,0739	2,4055	2,8188	3,1188
23	0,68531	1,1802	1,7139	2,0687	2,3979	2,8073	3,1040
24	0,68485	1,1789	1,7109	2,0639	2,3910	2,7969	3,0905
25	0,68443	1,1777	1,7081	2,0595	2,3846	2,7874	3,0782
26	0,68405	1,1766	1,7056	2,0555	2,3788	2,7787	3,0669
27	0,68370	1,1757	1,7033	2,0518	2,3734	2,7707	3,0565
28	0,68335	1,1748	1,7011	2,0484	2,3685	2,7633	3,0469
29	0,68304	1,1739	1,6991	2,0452	2,3638	2,7564	3,0380
30	0,68276	1,1731	1,6973	2,0423	2,3596	2,7500	3,0298
40	0,68066	1,1673	1,6839	2,0211	2,3289	2,7045	2,9712
60	0,67862	1,1616	1,6707	2,0003	2,2991	2,6603	2,9146
120	0,67656	1,1559	1,6577	1,9799	2,2699	2,6174	2,8599
∞	0,67449	1,1503	1,6449	1,9600	2,2414	2,5758	2,8070

Tabela 5. Dígitos aleatórios

03991	10461	93716	16894	98953	73231	39528	72484	82474	25593
38555	95554	32886	59780	09958	18065	81616	18711	53342	44276
17546	73704	92052	46215	15917	06253	07586	16120	82641	22820
32643	52861	95819	06831	19640	99413	90767	04235	13574	17200
69572	68777	39510	35905	85244	35159	40188	28193	29593	88627
24122	66591	27699	06494	03152	19121	34414	82157	86887	55087
61196	30231	92692	61773	22109	78508	63439	75363	44989	16822
30532	21704	10274	12202	94205	20380	67049	09070	93399	45547
03788	97599	75867	20717	82037	10268	79495	04146	52162	90286
48228	63379	85783	47619	87481	37220	91704	30552	04737	21031
88618	19161	41290	67312	74857	15957	48545	35247	18619	13674
71299	23853	05870	01119	92784	26340	75122	11724	74627	73707
27954	58909	82444	99005	04921	73701	92904	13141	32392	19763
80863	00514	20247	81759	45197	25332	69902	63742	78464	22501
33564	60780	48460	85558	15191	18782	94972	11598	62095	36787
90899	75754	60833	25983	01291	41349	19152	00023	12302	80783
78038	70267	43529	06318	38384	74761	36024	00867	76378	41605
55986	66485	88722	56736	66164	49431	94458	74284	05041	49807
87539	08823	94813	31900	54155	83436	54158	34243	46978	35482
16818	60311	74457	90561	72848	11834	75051	93029	47665	64382
34677	58300	74910	64345	19325	81540	60365	94653	35075	33949
45305	07521	61318	31855	14413	70951	83799	42402	56623	34442
59747	67277	76503	34513	39663	77544	32960	07405	36409	83232
16520	69676	11654	99893	02181	68161	19322	53845	57620	52606
68652	27376	92852	55866	88448	03584	11220	94747	07399	37408

(*continua*)

(*continuação*)

79375	95220	01159	63267	10622	48391	31751	57260	68980	05339
33521	26665	55823	47641	86225	31704	88492	99382	14454	04504
59589	49067	66821	41575	49767	04037	30934	47744	07481	83828
20554	91409	96277	48257	50816	97616	22888	48893	27499	98748
59404	72059	43947	51680	43852	59693	78212	16993	35902	91386
42614	29297	01918	28316	25163	01889	70014	15021	68971	11403
34994	41374	70071	14736	65251	07629	37329	33295	18477	65622
99385	41600	11133	07586	36815	43625	18637	37509	14707	93997
66497	68646	78138	66559	64397	11692	05327	82162	83745	22567
48509	23929	27482	45476	94515	25624	95096	67946	16930	33361
15470	48355	88651	22596	83761	60873	43253	84145	20368	07126
20094	98977	74843	93413	14387	06345	80854	09279	41196	37480
73788	06533	28597	20405	51321	92246	80088	77074	66919	31678
60530	45128	74022	84617	72472	00008	80890	18002	35352	54131
44372	15486	65741	14014	05466	55306	93128	18464	79982	68416
18611	19241	66083	24653	84609	58232	41849	84547	46850	52323
58319	15997	08355	60860	29735	47762	46352	33049	69248	93460
61199	67940	55121	29281	59076	07936	11087	96294	14013	31792
18627	90872	00911	98936	76355	93779	52701	08337	56303	87315
00441	58997	14060	40619	29549	69616	57275	36898	81304	48585
32624	68691	14845	46672	61958	77100	20857	73156	70284	24326
65961	73488	41839	55382	17267	70943	15633	84924	90415	93614
20288	34060	39685	23309	10061	68829	92694	48297	39904	02115
59362	95938	74416	53166	35208	33374	77613	19019	88152	00080
99782	93478	53152	67433	35663	52972	38688	32486	45134	63545

Tabela 6. Valores de $e^{-\mu}$

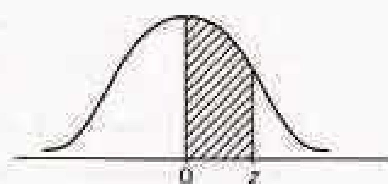
μ	0	1	2	3	4	5	6	7	8	9
0,0	1,0000	0,9900	0,9802	0,9704	0,9608	0,9512	0,9418	0,9324	0,9231	0,9139
0,1	0,9048	0,8958	0,8869	0,8781	0,8694	0,8607	0,8521	0,8437	0,8353	0,8270
0,2	0,8187	0,8106	0,8025	0,7945	0,7866	0,7788	0,7711	0,7634	0,7558	0,7483
0,3	0,7408	0,7334	0,7261	0,7189	0,7118	0,7047	0,6977	0,6907	0,6839	0,6771
0,4	0,6703	0,6636	0,6570	0,6505	0,6440	0,6376	0,6313	0,6250	0,6188	0,6126
0,5	0,6065	0,6005	0,5945	0,5886	0,5827	0,5770	0,5712	0,5655	0,5599	0,5543
0,6	0,5488	0,5434	0,5379	0,5326	0,5273	0,5220	0,5169	0,5117	0,5066	0,5016
0,7	0,4966	0,4916	0,4868	0,4819	0,4771	0,4724	0,4677	0,4630	0,4584	0,4538
0,8	0,4493	0,4449	0,4404	0,4360	0,4317	0,4274	0,4232	0,4190	0,4148	0,4107
0,9	0,4066	0,4025	0,3985	0,3946	0,3906	0,3867	0,3829	0,3791	0,3753	0,3716

μ	1	2	3	4	5	6	7	8	9	10
$e^{-\mu}$	0,36788	0,13534	0,04979	0,01832	0,006738	0,002479	0,000912	0,000335	0,000123	0,000045

Formato 17 x 24 cm
Papel Primapress 75 g/m² (miolo)
Supremo 250 g/m² (capa)
Número de páginas 320
Impressão Editora Santuário

Tabela 1. Áreas de uma distribuição normal padrão

Cada casa na tabela dá a proporção sob a curva inteira entre $z = 0$ e um valor positivo de z. As áreas para os valores de z negativos são obtidas por simetria.



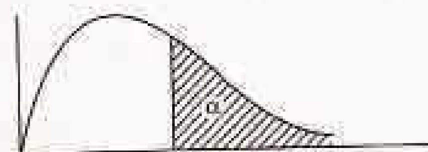
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4675	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936

(*continua*)

(*continuação*)

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Tabela 2. Distribuição de χ^2



φ \ α	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0001	0,0506	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	1,021	2,37	4,11	6,25	7,81	9,25	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,23	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8
16	5,14	5,80	6,91	7,96	8,31	11,9	15,3	19,4	23,5	26,3	28,4	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	53,4	35,7
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9;59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0

(*continua*)

(continuação)

φ \ α	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,5	42,8
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,2	33,1	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,5
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7

Para φ > 30 usar a aproximação: $\chi_x^2 = \frac{1}{2}\left[\pm Z_\alpha + \sqrt{2\varphi - 1}\right]^2$

Tabela 3. Distribuição de F de Snedecor α = 5%



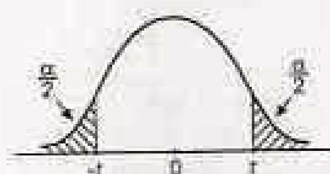
φ_2 \ φ_1	1	2	3	4	5	6	7	8	9	10	20	30	120	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	248,0	250,1	253,3	254,3
2	18.51	19.00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,45	19,46	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,66	8,62	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,80	5,75	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,56	4,50	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,87	3,81	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,44	3,38	3,27	2,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,15	3,08	2,97	2,92
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	2,94	2,86	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,77	2,70	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,65	2,57	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,90	2,85	2,80	2,75	2,54	2,47	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,46	2,38	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,39	3,31	2,18	2,13

(continua)

(continuação)

φ_2 \ φ_1	1	2	3	4	5	6	7	8	9	10	20	30	120	∞
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,33	2,25	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,28	2,19	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,23	2,15	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,19	2,11	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,16	2,07	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,12	2,04	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,10	2,01	1,87	1,81
22	4,30	3,44	4,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,07	1,98	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,05	1,96	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,03	1,94	1,79	1,73
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	1,93	1,84	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,21	2,18	2,12	2,08	1,84	1,74	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,75	1,65	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,66	1,55	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,57	1,46	1,22	1,00

Tabela 4. Distribuição *t* de Student



φ \ α	0,50	0,25	0,10	0,05	0, 025	0,01	0,005
1	1,00000	2,4142	6,3138	12.706	25.542	63.657	127,32
2	0,81650	1,6036	2,9200	4,3127	6,2053	9,9248	14,089
3	0,76489	1,4226	2,3534	3,1825	4,1765	5,8409	7,4533
4	0,74070	1,3444	2,1318	2,7764	3,4954	4,6041	5,5976
5	0,72669	1,3009	2,0150	2,5706	3,1634	4,0321	4,7733
6	0,71756	1,2733	1,9432	2,4469	2,9687	3,7074	4,3168
7	0,71114	1,2543	1,8946	2,3646	2,8412	3,4995	4,0293
8	0,70639	1,2403	1,8595	2,3060	2,7515	3,3554	3,8325
9	0,70272	1,2297	1,8331	2,2622	2,6850	3,2498	3,6897

(continua)

(continuação)

φ \ α	0,50	0,25	0,10	0,05	0, 025	0,01	0,005
10	0,69981	1,2213	1,8125	2,2281	2,6338	3,1693	3,5814
11	0,69745	1,2145	1,7959	2,2010	2,5931	3,1058	3,4966
12	0,69548	1,2089	1,7823	2,1788	2,5600	3,9545	3,4284
13	0,69384	1,2041	1,7709	2,1604	2,5326	3,0123	3,3725
14	0,692	1,2001	1,7613	2,1448	2,5096	2,9768	3,3257
15	0,69120	1,1967	1,7530	2,1315	2,4899	2,9467	3,2860
16	0,69013	1,1937	1,7459	2,1199	2,4729	2,9208	3,2520
17	0,68919	1,1910	1,7396	2,1098	2,4581	2,8982	3,2225
18	0,68837	1,1887	1,7341	2,1009	2,4450	2,8784	3,1966
19	0,68763	1,1866	1,7291	2,0930	2,4334	2,8609	3,1737
20	0,68696	1,1848	1,7247	2,0860	2,4231	2,8453	3,1534
21	0,68635	1,1831	1,7207	2,0796	2,4138	2,8314	3,1352
22	0,68580	1,1816	1,7171	2,0739	2,4055	2,8188	3,1188
23	0,68531	1,1802	1,7139	2,0687	2,3979	2,8073	3,1040
24	0,68485	1,1789	1,7109	2,0639	2,3910	2,7969	3,0905
25	0,68443	1,1777	1,7081	2,0595	2,3846	2,7874	3,0782
26	0,68405	1,1766	1,7056	2,0555	2,3788	2,7787	3,0669
27	0,68370	1,1757	1,7033	2,0518	2,3734	2,7707	3,0565
28	0,68335	1,1748	1,7011	2,0484	2,3685	2,7633	3,0469
29	0,68304	1,1739	1,6991	2,0452	2,3638	2,7564	3,0380
30	0,68276	1,1731	1,6973	2,0423	2,3596	2,7500	3,0298
40	0,68066	1,1673	1,6839	2,0211	2,3289	2,7045	2,9712
60	0,67862	1,1616	1,6707	2,0003	2,2991	2,6603	2,9146
120	0,67656	1,1559	1,6577	1,9799	2,2699	2,6174	2,8599
∞	0,67449	1,1503	1,6449	1,9600	2,2414	2,5758	2,8070

Tabela 5. Dígitos aleatórios

03991	10461	93716	16894	98953	73231	39528	72484	82474	25593
38555	95554	32886	59780	09958	18065	81616	18711	53342	44276
17546	73704	92052	46215	15917	06253	07586	16120	82641	22820
32643	52861	95819	06831	19640	99413	90767	04235	13574	17200
69572	68777	39510	35905	85244	35159	40188	28193	29593	88627

(continua)

(*continuação*)

24122	66591	27699	06494	03152	19121	34414	82157	86887	55087
61196	30231	92692	61773	22109	78508	63439	75363	44989	16822
30532	21704	10274	12202	94205	20380	67049	09070	93399	45547
03788	97599	75867	20717	82037	10268	79495	04146	52162	90286
48228	63379	85783	47619	87481	37220	91704	30552	04737	21031

88618	19161	41290	67312	74857	15957	48545	35247	18619	13674
71299	23853	05870	01119	92784	26340	75122	11724	74627	73707
27954	58909	82444	99005	04921	73701	92904	13141	32392	19763
80863	00514	20247	81759	45197	25332	69902	63742	78464	22501
33564	60780	48460	85558	15191	18782	94972	11598	62095	36787

90899	75754	60833	25983	01291	41349	19152	00023	12302	80783
78038	70267	43529	06318	38384	74761	36024	00867	76378	41605
55986	66485	88722	56736	66164	49431	94458	74284	05041	49807
87539	08823	94813	31900	54155	83436	54158	34243	46978	35482
16818	60311	74457	90561	72848	11834	75051	93029	47665	64382

34677	58300	74910	64345	19325	81540	60365	94653	35075	33949
45305	07521	61318	31855	14413	70951	83799	42402	56623	34442
59747	67277	76503	34513	39663	77544	32960	07405	36409	83232
16520	69676	11654	99893	02181	68161	19322	53845	57620	52606
68652	27376	92852	55866	88448	03584	11220	94747	07399	37408

79375	95220	01159	63267	10622	48391	31751	57260	68980	05339
33521	26665	55823	47641	86225	31704	88492	99382	14454	04504
59589	49067	66821	41575	49767	04037	30934	47744	07481	83828
20554	91409	96277	48257	50816	97616	22888	48893	27499	98748
59404	72059	43947	51680	43852	59693	78212	16993	35902	91386

42614	29297	01918	28316	25163	01889	70014	15021	68971	11403
34994	41374	70071	14736	65251	07629	37329	33295	18477	65622
99385	41600	11133	07586	36815	43625	18637	37509	14707	93997
66497	68646	78138	66559	64397	11692	05327	82162	83745	22567
48509	23929	27482	45476	94515	25624	95096	67946	16930	33361

(*continua*)

(*continuação*)

15470	48355	88651	22596	83761	60873	43253	84145	20368	07126
20094	98977	74843	93413	14387	06345	80854	09279	41196	37480
73788	06533	28597	20405	51321	92246	80088	77074	66919	31678
60530	45128	74022	84617	72472	00008	80890	18002	35352	54131
44372	15486	65741	14014	05466	55306	93128	18464	79982	68416
18611	19241	66083	24653	84609	58232	41849	84547	46850	52323
58319	15997	08355	60860	29735	47762	46352	33049	69248	93460
61199	67940	55121	29281	59076	07936	11087	96294	14013	31792
18627	90872	00911	98936	76355	93779	52701	08337	56303	87315
00441	58997	14060	40619	29549	69616	57275	36898	81304	48585
32624	68691	14845	46672	61958	77100	20857	73156	70284	24326
65961	73488	41839	55382	17267	70943	15633	84924	90415	93614
20288	34060	39685	23309	10061	68829	92694	48297	39904	02115
59362	95938	74416	53166	35208	33374	77613	19019	88152	00080
99782	93478	53152	67433	35663	52972	38688	32486	45134	63545

Tabela 6. Valores de $e^{-\mu}$

μ	0	1	2	3	4	5	6	7	8	9
0,0	1,0000	0,9900	0,9802	0,9704	0,9608	0,9512	0,9418	0,9324	0,9231	0,9139
0,1	0,9048	0,8958	0,8869	0,8781	0,8694	0,8607	0,8521	0,8437	0,8353	0,8270
0,2	0,8187	0,8106	0,8025	0,7945	0,7866	0,7788	0,7711	0,7634	0,7558	0,7483
0,3	0,7408	0,7334	0,7261	0,7189	0,7118	0,7047	0,6977	0,6907	0,6839	0,6771
0,4	0,6703	0,6636	0,6570	0,6505	0,6440	0,6376	0,6313	0,6250	0,6188	0,6126
0,5	0,6065	0,6005	0,5945	0,5886	0,5827	0,5770	0,5712	0,5655	0,5599	0,5543
0,6	0,5488	0,5434	0,5379	0,5326	0,5273	0,5220	0,5169	0,5117	0,5066	0,5016
0,7	0,4966	0,4916	0,4868	0,4819	0,4771	0,4724	0,4677	0,4630	0,4584	0,4538
0,8	0,4493	0,4449	0,4404	0,4360	0,4317	0,4274	0,4232	0,4190	0,4148	0,4107
0,9	0,4066	0,4025	0,3985	0,3946	0,3906	0,3867	0,3829	0,3791	0,3753	0,3716

μ	1	2	3	4	5	6	7	8	9	10
$e^{-\mu}$	0,36788	0,13534	0,04979	0,01832	0,006738	0,002479	0,000912	0,000335	0,000123	0,000045

CURSO DE ESTATÍSTICA

Este texto reúne matéria essencial para cursos universitários de Estatística, centrando-se no estudo das probabilidades, inferência e estatística não-paramétrica.

Fruto de experiências vivenciadas na condução de cursos de Estatística para o 3º grau nas áreas de exatas e humanas, o livro traz farta exemplificação de aplicações do cálculo das probabilidades, modelos de distribuições de probabilidades, inferência estatística, técnicas não paramétricas e análise da variância.

Esta nova versão, após 15 anos de sua primeira edição, tem características eminentemente didáticas: detalhada exposição, solução de exemplos e proposição de exercícios cujas respostas se encontram no final do livro.

Os quatro primeiros capítulos são dedicados ao estudo do cálculo das probabilidades. O capítulo 5 expõe, detalhadamente, a estatística descritiva. Os capítulos 6, 7, 8 e 9 apresentam as distribuições amostrais; amostragem, intervalos de confiança e teste de hipóteses. As técnicas não paramétricas são destacadas no capítulo 10 e a análise da variância constitui o último capítulo.

NOTA SOBRE OS AUTORES

JAIRO SIMON DA FONSECA é professor titular do Departamento de Administração da FEA/USP. É co-autor de *Estatística aplicada e Marketing,* publicados pela Atlas.

GILBERTO DE ANDRADE MARTINS é doutor em Administração pela FEA/USP. É professor do Departamento de Administração da FEA/USP, Fundação Santo André e IMES-SCS. É autor de *Manual para elaboração de monografias* e co-autor de *Estatística aplicada* e *Princípios de estatística*, publicados pela Atlas.

APLICAÇÃO

Livro-texto para a disciplina ESTATÍSTICA dos cursos de graduação em Administração de Empresas, Economia, Contabilidade, Matemática e Engenharia.

publicação atlas

